SCIENTIFIC AMERICAN

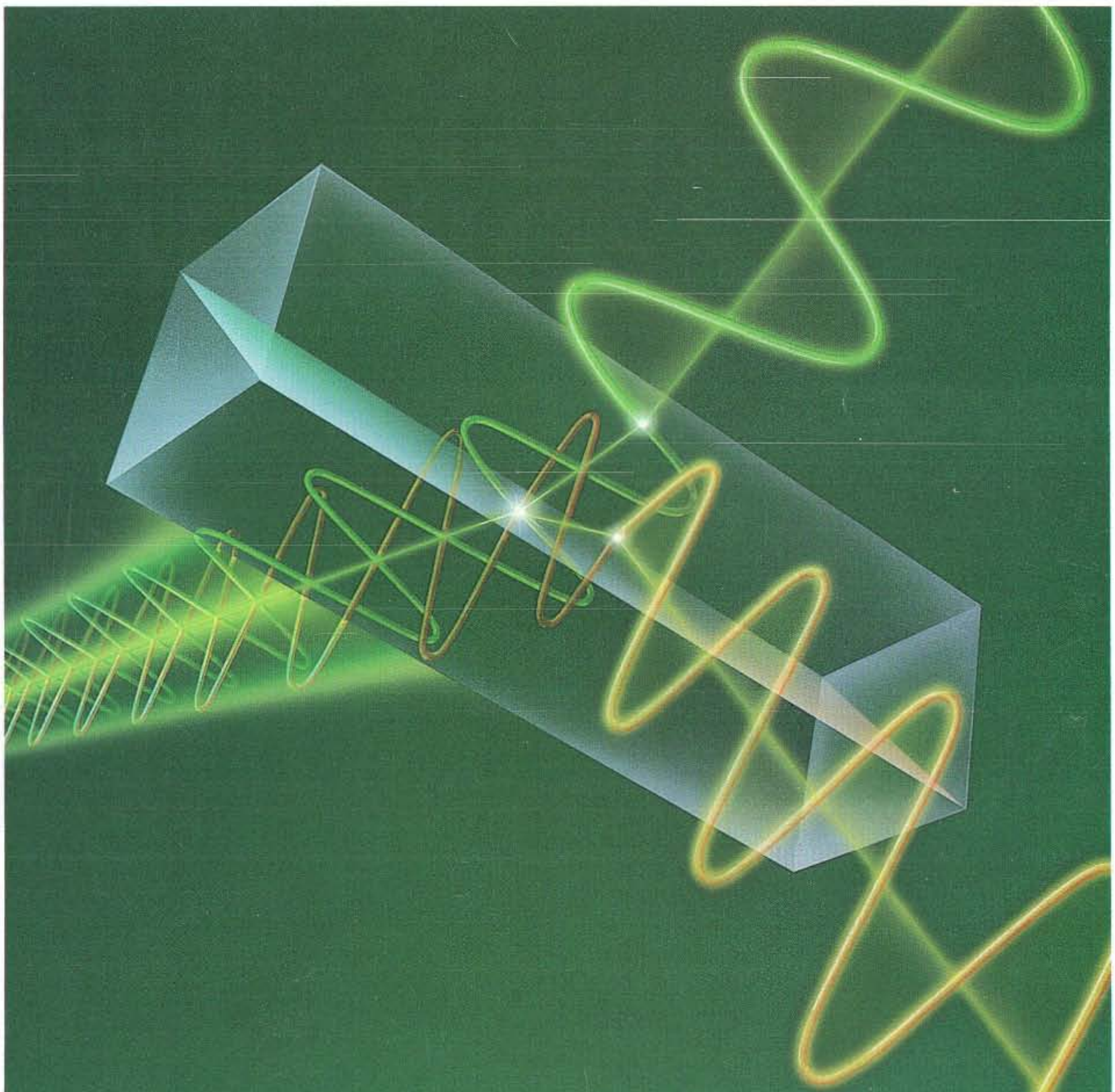
OCTOBER 1992

$3.95

The promise of diamond semiconductors.

Was early man a heroic hunter—or a scavenger?

Raising the grades in U.S. science education.

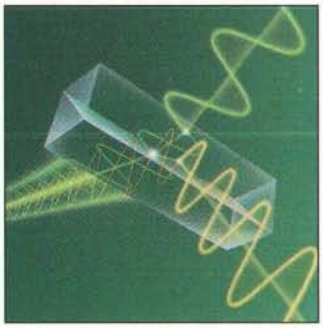


Light signals *split by a simple prism allow messages to be transmitted in absolute secrecy.*

SCIENTIFIC AMERICAN

October 1992 Volume 267 Number 4

DEPARTMENTS



THE COVER painting features a calcite Wollaston prism, the key component in a cryptographic system that conveys information in absolute secrecy. The prism shifts photons polarized in the horizontal direction to the left, while directing vertically polarized photons toward the right. By generating and then measuring photons with different polarization, two people can send messages and detect eavesdropping (see "Quantum Cryptography," by Charles H. Bennett, Gilles Brassard and Artur K. Ekert, page 26).

THE ILLUSTRATIONS

Cover painting by Tomo Narashima

Page	Source
18–19	*The Terrestrial Paradise,* by Jan Brueghel the Elder, Museo Laraz Galdiano/Art Resource
20	Bettmann Archive (*top*), Jana Brenning (*bottom*)
21	Jana Brenning
22	Johnny Johnson (*graph*), Jana Brenning (*drawing*)
23	Guilbert Gates/JSD
24	Kathleen Katims/JSD
27	Robert Prochnow
28	Courtesy of David Kahn, © 1983 Macmillan Publishing Company
29–33	Michael Goodman
35	Galen Rowell/Mountain Light Photography
36	Patricia J. Wynne (*top*), Carol Donner (*bottom*)
37–38	Patricia J. Wynne
41	Timothy J. Richmond, Library of the Swiss Federal Institute of Technology, Zurich
42	Ian Worpole; Barbara A. Hamkalo, University of California, Irvine (*insets*)
44–45	Ian Worpole
46	Stephen J. Kron and Gerald R. Fink, Whitehead Institute (*top*), Ian Worpole (*bottom*)
47	Ian Worpole
57–59	Philip J. DeVries
60	Philip J. DeVries (*left*), Jason Goltz (*right*)
61	Patricia J. Wynne
62	Johnny Johnson (*diagram*), Patricia J. Wynne (*drawings*)
65	Jared Schneidman
66	Robert Clausing, Oak Ridge National Laboratory
67	Jared Schneidman
68	Guilbert Gates/JSD (*top*), Jared Schneidman (*bottom*)
69	Guilbert Gates/JSD
70–71	Patricia J. Wynne
72	Robert J. Blumenschine
73	Leonard Lee Rue/Bruce Coleman, Inc. (*photograph*), Patricia J. Wynne (*drawings*)
74	Patricia J. Wynne
75	Johnny Johnson (*graph*), Patricia J. Wynne (*drawings*)
76	John A. Cavallo
78–79	Richard Nowitz, © National Geographic Society
80	Johnny Johnson
82	Randy Santos/Randolf Photography
83	Robert Prochnow
84–85	Lois S. Kim, St. Andrew's School (*left*), Max Aguillera-Hellweg (*right*)
86	Randy Santos/Randolf Photography
95–97	Jana Brenning

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Alan Hall, *Executive Editor;* Michelle Press, *Managing Editor;* Timothy M. Beardsley; Elizabeth Corcoran; Deborah Erickson; Marguerite Holloway; John Horgan, *Senior Writer;* Philip Morrison, *Book Editor;* Corey S. Powell; John Rennie; Philip E. Ross; Ricki L. Rusting; Russell Ruthen; Gary Stix; Paul Wallich; Philip M. Yam

ART: Joan Starwood, *Art Director;* Edward Bell, *Art Director, Graphics Systems;* Jessie Nathans, *Associate Art Director;* Nisa Geller, *Photography Editor;* Johnny Johnson

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Daniel C. Schlenoff

PRODUCTION: Richard Sasso, *Vice President Production & Distribution; Managers:* Carol Albert, *Prepress;* Tanya DeSilva, *Projects;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Leo J. Petruzzi, *Manufacturing & Makeup;* William Sherman, *Quality Assurance;* Carl Cherebin

CIRCULATION: Lorraine Leib Terlecki, *Circulation Director;* Cary Zel, *Circulation Manager;* Rosa Davis, *Fulfillment Manager;* Katherine Robold, *Newsstand Manager*

ADVERTISING: Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: Michelle Larsen, *Director, New Business Development;* William Buchanan; Peter Fisch; Meryle Lowenthal; Elizabeth Ryan. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager;* William F. Moore. WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Kate Dobson, *Advertising Manager;* Lisa K. Carden, Lianne Bloomer, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: EUROPE: GWP International, Düsseldorf; Roy Edwards, GWP International, London; Kurt Bühlmann, GWP International, Zürich. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd. SPECIAL PROJECTS: Barth David Schwartz, *Director*

ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie M. Beaumonte, *Business Manager*

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550

PRESIDENT AND CHIEF EXECUTIVE OFFICER: John J. Hanley

CORPORATE OFFICERS: *Executive Vice President and Chief Financial Officer,* R. Vincent Barger; *Senior Vice President,* Linda Chaput; *Vice Presidents:* Jonathan Piel, John J. Moeling, Jr.

CHAIRMAN OF THE BOARD: Dr. Pierre Gerckens

CHAIRMAN EMERITUS: Gerard Piel

LETTERS TO THE EDITORS

On Treating AIDS

As a professional concerned with the planning and management of AIDS treatment programs, I read with interest "Understanding the AIDS Pandemic," by Roy M. Anderson and Robert M. May [SCIENTIFIC AMERICAN, May]. The authors demonstrate the danger of moving too hastily even to provisional conclusions when they suggest that zidovudine (AZT) treatment to slow the progress of the disease may be beneficial to the individual but not necessarily for the community. Among the side effects of AZT are persistent nausea and other physical conditions that affect sexual activity. Anecdotal reports suggest that patients far enough along in the course of the disease to be on AZT tend to be concerned about how to restore their sexual drives. I am led to the conclusion that AZT therapy does not raise the risk of transmission and may actually decrease it.

My concern is that in an otherwise informative and useful article, even the tentative suggestion that AZT therapy may be detrimental to the public will influence some physicians who are either uninformed or prejudiced to withhold life-prolonging treatments from individuals who do not represent a public health risk. It is not farfetched to expect that uninformed policymakers will seize on such statements and use them to limit the support for treatments.

MENACHEM EMANUEL
Chicago, Ill.

Anderson and May reply:
We most certainly would not wish the questions raised in our article to be misused to argue for withholding AZT treatments or the funds for them. We do, however, continue to believe that considerations of transmissibility should, as far as possible, play a larger and more explicit part in the protocols of drug development than they currently do.

Recalculating Camels

I enjoyed the generalization of the old Arabian inheritance problem in Ian Stewart's "Mathematical Recreations" column [SCIENTIFIC AMERICAN, June] and immediately tried my hand at it. The 14 presented solutions to the equation $(1/a + 1/b + 1/c + 1/d = 1)$ are correct. The original problem, however, requires that d be evenly divisible by a, b and c to prevent the sons from receiving fractional camels. Two of the solutions fail this condition: $a = 2$, $b = 3$, $c = 10$ and $d = 15$; and $a = 3$, $b = 4$, $c = 4$ and $d = 6$.

I am also not sure that Stewart's generalization is in the spirit of the original problem, which emphasizes unequal inheritances. One of his solutions, in which all the variables equal four, is equivalent to a parental will that says, "Let each of my three sons have one fourth of our three camels." If the three sons need a wise man with an extra camel for that problem, they deserve to eat ground camel meat.

JAMES J. MCHUGH
Chicago, Ill.

Phantom Pain

If neurologist S. Weir Mitchell wanted to test his colleagues' reaction to the idea of phantom limbs by publishing a fictional tale about it, he need not have bothered ["Phantom Limbs," by Ronald Melzack, SCIENTIFIC AMERICAN, April]. In 1851, 15 years before his story was published, the most famous amputee in literature had already spoken on the issue. In chapter 108 of *Moby Dick*, Captain Ahab says, "Look, put thy live leg here in the place where mine once was; so, now, here is only one distinct leg to the eye yet two to the soul. Where thou feelest tingling life; there, exactly there, there to a hair, do I.... I still feel the smart of my crushed leg, though it be now so long dissolved."

ROB HARDY
U.S. Air Force

Not so Hottentot

In "Eloquent Remains" [SCIENTIFIC AMERICAN, May], Philip E. Ross refers to my work, in which I have pointed out that the relatively high percentage of lactose tolerance in some San populations supports the view that the San have often oscillated between different modes of subsistence or else have repeatedly been in close contact with pastoral peoples. There is no basis for Ross's statement that the San-speaking foragers "split off from neighboring Hottentots renouncing the pastoral way of life the Hottentots follow to this day." It cannot be deduced from what I wrote and is pure fiction.

MICHAEL J. CASIMIR
University of Cologne

Ross replies:
Casimir may misunderstand what I meant by "split off." Most experts agree the two peoples' languages and genotypes are closely related.

No Sparking

"Electric Car Pool," by Gary Stix ["Science and Business," SCIENTIFIC AMERICAN, May], rightly characterized the technical goals of the battery consortium as highly ambitious, but it glossed over some serious problems. First, electric vehicles are not pollution free; they simply move the atmospheric emissions from the vehicle to other—one hopes, distant—places. Although electric power plants can achieve higher thermal efficiencies than current gasoline engines, the losses in long-distance power transmission, battery charging and discharging, and vehicle motor control raise the total consumption of primary fuels and, correspondingly, atmospheric pollution.

The vehicle manufacturers are willing to support the battery development program only with substantial taxpayer funding. Support from the electric utilities is based on the benefits of higher energy sales at night, when batteries will recharge. Those benefits will be negligible, however, until electric vehicles are driving a significant fraction of the distance covered by all vehicles. If this fraction ever reaches about 25 percent, the utilities will probably have difficulties meeting the night peak loads.

Your photograph of the General Motors Impact vividly illustrates the problem of installing a ton of lead in an otherwise lightweight vehicle and justifies the name for this model! The consequences of even a low-speed impact could be serious with lead-acid batteries and would almost certainly be worse with liquid sodium/sulfur batteries.

GEOFFREY J. CRASK
Santa Ana, Calif.

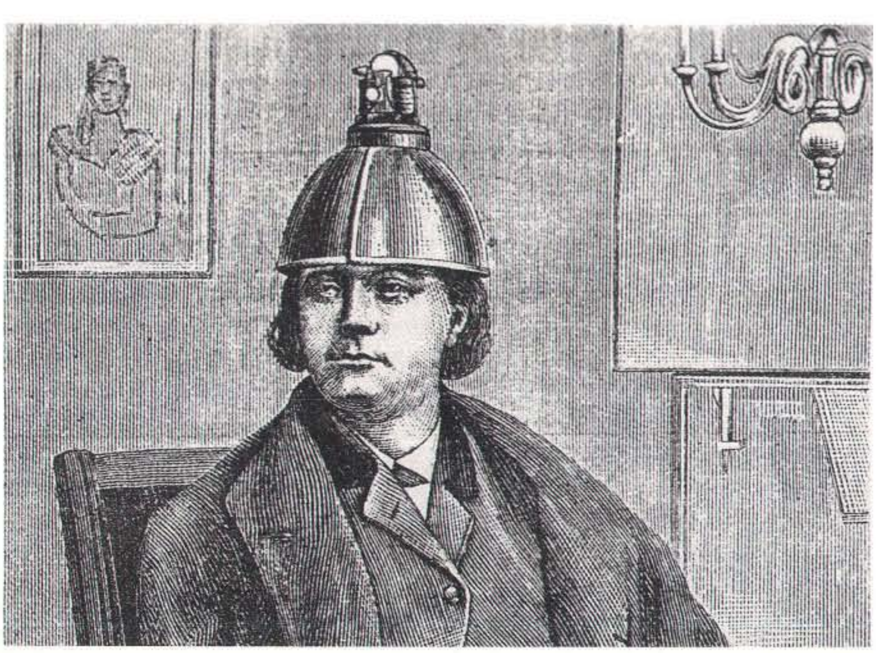
50 AND 100 YEARS AGO

OCTOBER 1942

"'The problems we are going to meet after this war is over are two-fold,' said Mr. James F. Lincoln, President of the Lincoln Electric Company. 'First of all, there will be great unemployment because even if the call for goods is so great to keep all men employed, the reshuffling of these men to peace-time work will take time. The second thing is that competition will have changed due to the existence of government financed organizations which will have less overhead than those who have financed their own expansions. Of course, there are many unknown features. For instance, there is no way of knowing what kind of government we will have. It is rather obvious from history that it will be totalitarian. No bankrupt nation ever escaped totalitarianism. Perhaps all of our planning will go for naught because we will be told by bureaucrats what to do and how to do it.'"

"Miss M. F., aged 25, had had symptoms of schizophrenia for about ten years but was first frankly diagnosed in April 1939. Neither insulin shocks nor metrazol altered her state. On July 29, 1940, general hypothermia was tried. Her temperature was kept below 98 degrees for 29 hours, and below 95 for 22 hours, at one time reaching a minimum of 83. 'Following hypothermia she returned to her home,' report Drs. John H. Talbott and Kenneth J. Tillotson of Harvard Medical School. 'At the last follow-up report it was apparent that she had made a good social recovery and a satisfactory social adjustment.'"

"The business of smashing atoms to release great gusts of energy is a profitable sport—for news reporters. But it is not an item that has much standing in physics laboratories. Radioactive materials, of which there are only minute amounts in the earth, disintegrate and slowly release large amounts of energy. If radium, for example, were as plentiful as copper, atomic boilers using radium as fuel might be practical, but there just is not very much radium available. As far as artificial disintegrations are concerned, the verdict thus far seems to be definitely thumbs down on such operations for giving a net yield of energy. Far more energy has to be put into the operation than can be got out. There is some evidence that one of the isotopes of uranium, if relatively pure, might, upon bombardment with neutrons, disintegrate to give a net yield of a rather large amount of energy. But this isotope of uranium is one of the rarest of rare materials."



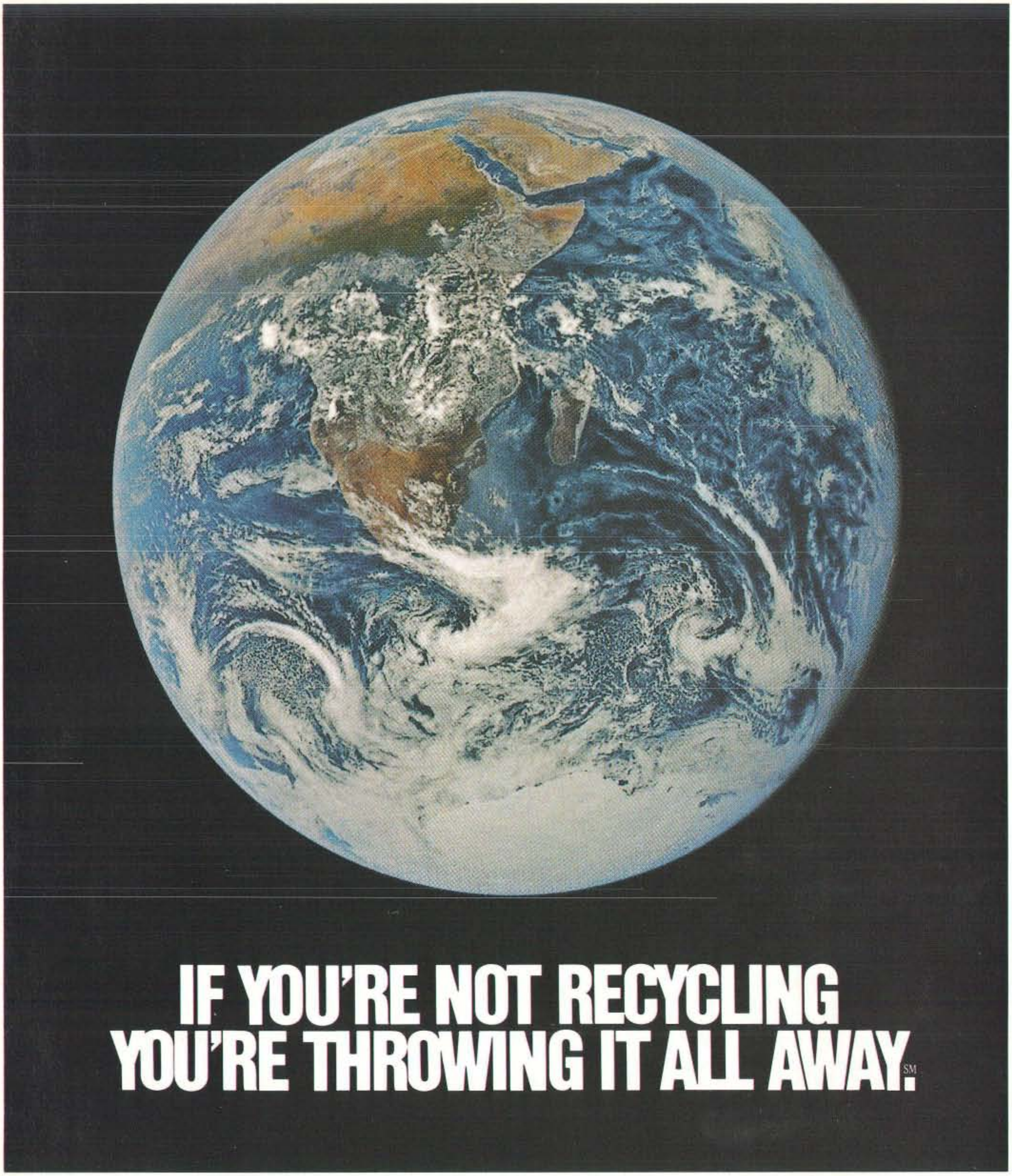
Tourette's vibrating helmet

SCIENTIFIC AMERICAN

OCTOBER 1892

"The perfection of the science of long distance telephony has been going on for the past five or six years, until an epoch of much interest has finally been reached; that is the perfect transmission of articulate speech for a distance of one thousand miles and over. We were invited to attend the first public demonstration of this fact on the afternoon of October 18, at the main offices of the Long Distance Division of the American Telephone and Telegraph Company in New York City. It should be mentioned that an important element in the success of long distance telephony is the improved battery now used for energizing the transmitter, which has the merit of maintaining a nearly uniform electro-motive force of high tension for an extensive period of time."

"The trembling palsy, also called Parkinson's disease, is a sort of painful nervous disorder that deprives the unfortunate who is afflicted with it of rest and sleep. Professor Charcot a long time ago learned from some invalids who were troubled with this infirmity that they derived decided relief from long rides on a railroad or in a carriage. So Mr. Charcot had an arm chair constructed to which a to and fro motion was given by means of an electrical windlass. There could be nothing more insupportable for a well person than such shocks, which demolish you, put you out of order, and shake up your intestines, and after half a minute's experience you would ask for mercy. The invalid, on the contrary, lolls in the chair as you would do on a soft sofa. The more he is shaken the better he feels. The method was some time ago singularly improved by Dr. Gilles de la Tourette, a pupil of Mr. Charcot. He had an apparatus constructed for the treatment of megrims and nervous headaches; it was the vibrating helmet [*see illustration at left*]. Upon this helmet there is a small alternating current motor that makes about 600 revolutions per minute. In a few minutes a sort of general lassitude is experienced, with a tendency to sleep.—*La Nature.*"

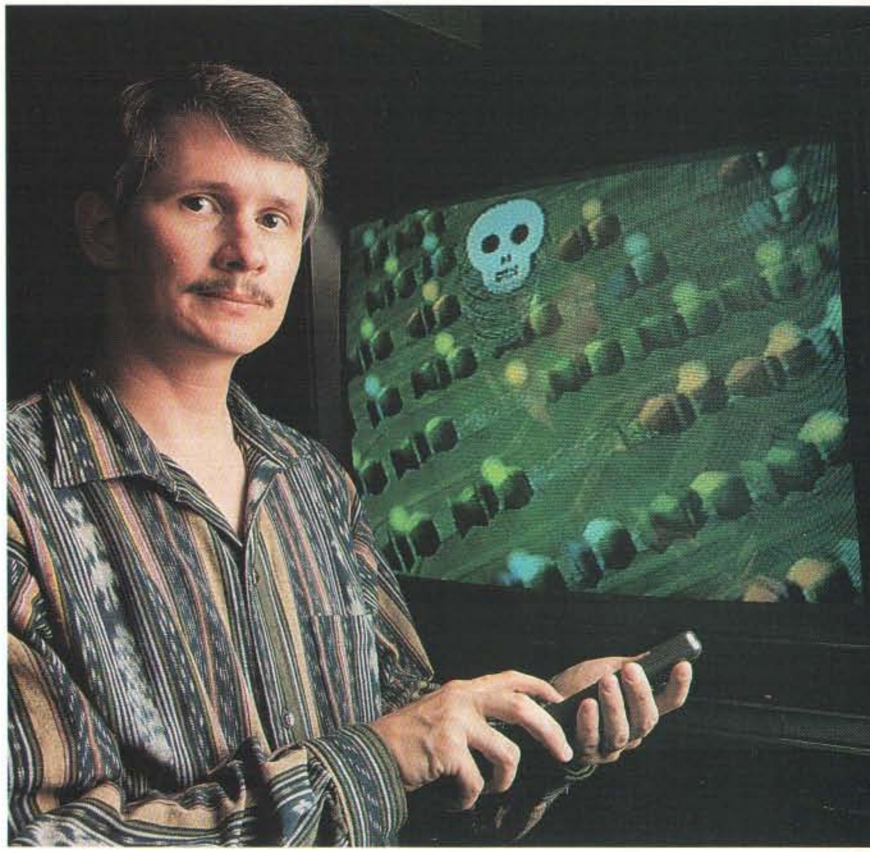


A little reminder from the Environmental Defense Fund that if you're not recycling, you're throwing away a lot more than just your trash.

You and your community can recycle. Please write the Environmental Defense Fund at: EDF-Recycling, 257 Park Avenue South, New York, NY 10010, for a free brochure that will tell you virtually everything you need to know about recycling.

The Edge of Chaos

Complexity is a metaphor at the Santa Fe Institute



COMMUNITY OF DIGITAL ORGANISMS, created by Thomas S. Ray, begins with single-cell "ancestors," then evolves parasites. A death's-head claims the old and defective. Computer animation: Anti-Gravity Workshop. Photo: Chel Beeson.

Just to the left of the viewgraph projector, Leo Buss was fidgeting. He was on deck, the next speaker at the Santa Fe Institute's "integrative themes" workshop held this past summer. Around the conference table before him sat physics Nobel laureates Murray Gell-Mann and Philip W. Anderson, thumbing through newspapers and doodling on notepads. Another two dozen or so researchers—biologists, physicists, economists, computer scientists, mathematicians—tried to lounge in the stiff-backed chairs.

"I have a long-standing frustration with evolutionary theory," Buss began, pacing tight arcs in the corner. "Evolution is preoccupied with frequency—the frequency of alleles, of individuals," the Yale University biologist growled. "That's okay if you have classes of organisms and know the relationships between them. But it doesn't say how you got classes in the first place."

That kind of passion—frustration with orthodoxy and eagerness to chip away at science's bedrock assumptions—unites and ignites discussions at the Santa Fe Institute. For the past six years, this New Mexico research retreat has been a simmering cauldron of Big Questions: How did the world become so complicated? Why is there so much order and structure in the world—and simultaneously so much instability? Why does innovation seem to thrive at the boundary separating order and disorder? Are there common patterns in how systems—be they economic, ecological, biological or something else—adapt and change?

The answers, believe the researchers who frequent the institute, lie in a rich collection of patterns that occur again and again in systems as diverse as the economy of Silicon Valley, the ecosystem of a pond or the web of algorithms running on a computer. The scientists describe these and other changing communities as "complex adaptive systems." By weaving together insights from many disciplines, these scientists hope to create a broad understanding of how things come into being and evolve.

Anywhere else such an undertaking might be considered an evening parlor game. "It's not the way I do science," concedes John Maynard Smith, an evolutionary biologist at the University of Sussex in Brighton, England. "I've never been able to work from the general to a particular field." Yet he found the summer workshop intriguing and admits he learned a new approach or two to studying the dynamics of a system. "They're extremely clever people," he adds. "I hope they succeed."

Indisputably, the institute has built a reputation for doing the unconventional. It is the ongoing host of "artificial life" workshops, a Woodstock for computer hackers. Artificial life mimics and models simple living systems with computer algorithms and occasionally a motor and set of wheels. Among the great hits at the third such conference, held this past spring, was an artificial ecosystem called Tierra, devised by biologist Thomas S. Ray of the University of Delaware. As single-cell "organisms" in the form of computer code begin to replicate in Tierra, they mutate and compete with one another for resources. Parasites spontaneously arise, followed by new parasites that feed on the older varieties. With Tierra, Ray can watch complex communities and novelty emerge.

For all its big goals, the institute operates on a sometimes wobbly budget, $3 million this year. One third of the funding is provided by government sources, another third comes from foundations and the rest is from corporations and individuals. Although some 50 visiting scientists regularly participate in workshops and contribute working papers, only a handful have spent more than a few weeks at a time at Santa Fe.

The institute has wielded its interdisciplinary approach with skill, in part because most of the researchers share a comfortable familiarity with sophisticated mathematics and computer modeling. "There's a tendency for dilettantism to enter into interdisciplinary proj-

ects," observes William A. Brock, an economist at the University of Wisconsin and a Santa Fe regular. In contrast, "the institute has forced us to weld together to do something that's not trivial," Brock says. "That multidisciplinary role gives the institute a lot more standing," adds Eric Bloch, formerly head of the National Science Foundation and now a distinguished fellow at the Council on Competitiveness in Washington, D.C. "It's a unique approach that's difficult to duplicate in other institutes."

Nonlinear systems formed the thread that first pulled together the diverse scholars at the Santa Fe Institute. Nonlinear systems are ones in which the shape of the whole is not easily predicted by looking at the parts. But underlying even the seeming randomness of chaos are a small number of characteristic, albeit complicated, patterns of behavior, called attractors. "When it started, Santa Fe was the only organized group of researchers seriously thinking about chaos and nonlinear systems," Bloch observes. "That's no longer the

The Best Little Yard Sale in New Mexico

If memory serves him correctly, Edward B. Grothus believes that the six-foot graphite spheres wound up decorating someone's yard on the chichi side of Santa Fe. "They were beautiful—all graphite with something in the middle," he recalls. "I called all over the lab trying to find out what in the world they made these things for. Never found out."

That's not so surprising. The "lab" was none other than that New Mexico fixture, the Los Alamos National Laboratory, home of the Manhattan Project and the first nuclear weapons. For more than 20 years, Grothus has been one of the most avid buyers of the secretive installation's technological castoffs. In turn, he offers the choice bits of surplus to others through his own scrapyard, the Los Alamos Sales Company. "You hear terrible things about waste in society," Grothus muses. "Los Alamos, to its credit, made a lot of things available to the general public."

Getting rid of obsolete equipment is no small task, particularly at a place determined to be on the cutting edge. Since the earliest days of the lab, Los Alamos managers have tried to recycle old equipment, first by offering it to other government groups or schools and then by trying to sell the rest to the public at (very significant) discounts. A metal office desk, for instance, might go for $15.

As a result, Los Alamos's yard sales became locally famous. "About 125 loyal people attend every week," estimates Frank J. Morbillo, a sculptor at the Shidoni Foundry and Galleries, five miles north of Santa Fe. Prospective buyers take up positions near the roped-off grounds on Thursdays at noon. After a brief safety lecture, the head of the yard unleashes the crowds. "He pretty much says, 'On your mark, get set, go!'" Morbillo recounts.

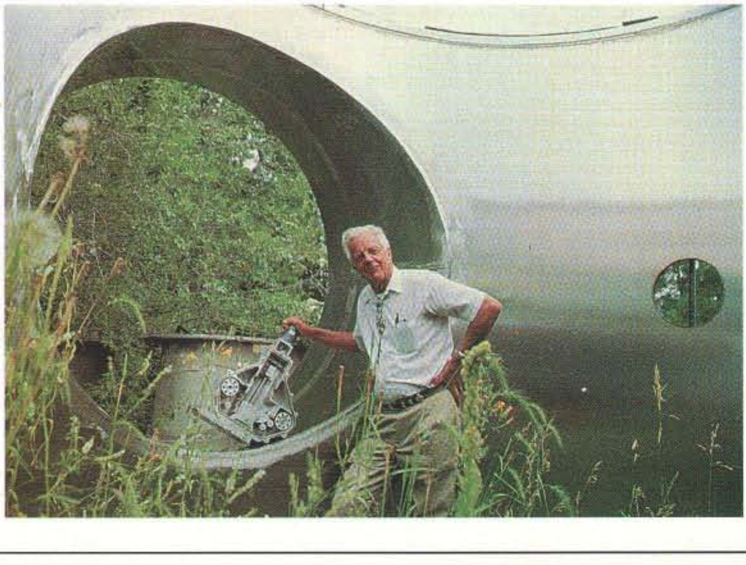
It goes without saying that radioactive, hazardous and classified materials are excluded from such sales. But anything else—oscilloscopes, computers, mobile offices and desk chairs—may find its way into the public domain. Since last December, Los Alamos has shipped expensive items to the Bentley Auction Service in Albuquerque. The lab continues to handle less glamorous materials itself.

Over the years, careers have been built on capturing the irony of the Los Alamos remains. T. A. ("Tony") Price, for one, has spent the past 28 years fashioning bits of remaindered metal into sculptures and masks that echo religious themes. Most of his works are now displayed at Biosphere 2, the experimental environment near Tucson, Ariz. "In the early days, a lot of the pieces, like hydrogen bomb casings, made beautiful sounds" when struck, Price recalls. Such parts are hard to come by these days, "but I still have a little stockpile left," he says.

Grothus, who worked for 20 years at Los Alamos as a mechanical engineer before he quit in 1969 as a protest against the Vietnam War, says he has tried to be discriminating about his choices. When the Reagan administration cut back solar-energy research, he snapped up the pieces of megawatt solar-cell systems. Remnants of now antique Cray supercomputers make regular appearances. "Whole disk-drive units come out," Grothus says. "Of course, about all we've ever sold are the recording heads." His most lucrative purchase: 120 tons of stainless-steel tubing, which he in turn happily sold to a Chicago steel company.

The operation has proved handy for Los Alamos, too. When the lab dismantled the Antares laser fusion project about seven years ago, Grothus acquired more than two dozen six-foot-long stainless-steel tanks. About a year ago lab managers found themselves needing a few such tanks—and so grudgingly bought back several. "They're mad because they have to pay me," Grothus chuckles, "but experiments come and go, and the lab just doesn't have room to keep all this stuff." (He did, nonetheless, donate back to the lab a tank called Herr Auge, which, according to Los Alamos lore, was "the eye that first saw the neutrino.")

Still, even Grothus is feeling cramped. Many items are piled on his lawn; his neighbors plan to take him to court this fall. "I haven't been to the Los Alamos salvage yard but once in the last month," he says. "I've got enough stashed away for rainy days. Now I'm just trying to placate the neighbors." *—Elizabeth Corcoran*



PIECES OF ANTARES laser fusion project are among Edward B. Grothus's wares. Photo: Chel Beeson.

case, but the Santa Fe Institute had a pioneering role."

More recently, the institute's attention has focused on the knottier questions posed by what the researchers call complexity and complex adaptive systems. Whereas chaotic systems have few variables and converge to a small number of attractors, "the systems we're interested in have an infinite number of attractors—or call it, choices of behavior—that they can choose to adopt" and countless variables, explains Anderson of Princeton University.

Defining complexity more tightly becomes a problem in its own right. Some systems may be complex, but they do not adapt. Per Bak, a senior scientist at Brookhaven National Laboratory, describes the tendency of large systems to drive themselves to a point where statistically predictable avalanches of change ensue as "self-organized criticality." (Sand piles illustrate this idea nicely.) Others picturesquely refer to criticality as "the edge of chaos," the border between order and disorder.

Other complex systems are composed of agents that are continuously adapting. Although simple, they gather information about the environment around them, sifting out relevant details from random noise. The agents then compress that information into models or schemata, which they use to anticipate and react to changes in their environment. Over time, they modify those schemata to reflect new information.

Yet how do adaptive agents evolve? According to a scenario proposed by Stuart A. Kauffman, a biochemist at the University of Pennsylvania School of Medicine, self-organized criticality occurs in models of coevolution. For example, communities with a subcritical or low diversity of organisms lack the momentum to develop explosively into a new breed; supracritical communities expand so rapidly that they consume all their available food and die off. At the edge of chaos, however, mutation and innovation occur, he argues.

Buss and Santa Fe postdoctoral fellow Walter Fontana take a different tack on the issue of how new classes of organisms arise. If the clock were set back to the beginning of time, what conditions would have to be present to produce contemporary classes of organisms, such as multicellular organisms? What other conditions have arisen in the course of development that shaped the specific paths—say, the rise of a frog? By trying to model mathematically the simplest interactions in a community, they hope to understand better "not the survival of the fittest, but the origin of the fittest," Buss says.

Curiously enough, one of the most apparent examples of a complex adaptive system has proved to be the hardest for the Santa Fe researchers to grapple with: the economy. There are ongoing efforts to model how stock-market players learn, as well as how cities change. "The character of our business is changing, and the dynamics aren't well described by classical economics," says Henry A. Lichstein, a vice president at Citicorp in New York City, which provides more than $100,000 a year to the institute.

The institute researchers argue that

Lying by the Book

Dissembling by public officials is probably as old as government. Certainly the practice has become a tradition in Washington. During the 1950s and 1960s, for example, U.S. Army planes carried out mock biological warfare attacks against American and Canadian cities by spraying them with live—though supposedly harmless—bacteria. If local officials asked what was going on, the army said it was testing a radar-deflecting chaff.

One might think that such prevarication—whether justified or not—is done on an ad hoc, seat-of-the-pants basis. That might have been the case previously, but no more. The Bush administration has actually drafted regulations on the use of deception to provide cover for secret programs. Bureaucrats' passion for secrecy, it seems, is exceeded only by their passion for codification.

The regulations are part of the National Industrial Security Program Operating Manual, which sets forth security procedures for government agencies and contractors involved with classified programs. Recently the Department of Defense generated a supplement to the manual for "special access" (also called "black") programs, whose existence cannot even be acknowledged. Dated May 29, 1992, and stamped "draft," the supplement states:

> Cover stories may be established for unacknowledged programs in order to protect the integrity of the program from individuals who do not have a need to know. Cover stories must be believable and cannot reveal any information regarding the true nature of the contract. Cover stories for Special Access Programs must have the approval of the PSO [Program Security Officer] prior to dissemination.

The supplement also notes that special access programs must have "nonattributable" telephone lines, also called "Hello lines," connecting them to the outside world. Personnel who answer such a telephone must "state the proper salutation, e.g. Good Morning or Hello. Do not use the company name."

Steven Aftergood of the Federation of American Scientists, which made the supplement public, professes to be shocked at the cover-story policy, which he calls "officially sanctioned lying." "One can see situations where this might be warranted, maybe in the midst of wartime," he says. "But this is not sufficiently well defined to convince me that it is limited. It's obviously a very dangerous practice, because it can corrupt the public discourse."

Susan Hansen, a spokesperson for the Pentagon, grumbles that the document on cover stories was confidential. "Whoever sent it to you was unauthorized," she says. She points out, furthermore, that the document is an unapproved draft version that "does not represent the policy of the federal government."

But does this statement itself represent a cover story? According to a Senate staff member specializing in security issues, the Bush administration has already implemented the cover-story policy—with the complicity of some congressional oversight committees. Indeed, the administration has consulted with Congress before disseminating cover stories about several "major programs" to the media, the staffer says. Such as? "Sorry, I can't tell you that," he replies.

The staff member emphasizes that Congress, although it gives its approval to cover stories when the need for security seems clear, does not actively participate in the deception. Indeed, Congress is trying to reduce the need for such deception by cutting the number of black programs to a minimum. "It's a very uncomfortable situation in the democratic framework to lie about what you're doing," he acknowledges. *—John Horgan*

their six years have been well spent. "I'm very excited about the feeling that there is a growing synthesis," declares Gell-Mann of the California Institute of Technology. "We're not saying there is a science of 'complexity,'" adds David Pines, a physicist at the University of Illinois. Instead "we have come up with a whole set of candidate metaphors for complex adaptive systems that look very promising for a broad range of systems," he says.

Science, suggests W. Brian Arthur, an economist at Stanford University, "is about the creation of metaphor." Before the 1680s, "people looked at the world and said it was messy, alive, organic," he says. "Fifty years after Newton, all people could see was order—order, stasis, equilibrium and harmony." Today tidy metaphors are insufficient. Emboldened by computing power and the metaphors of complexity, the researchers aim to "go beyond order and construct systems from simple premises that turn out to be very, very messy," Arthur says.

Over the next few years, Santa Fe researchers must breathe life into the metaphors they have devised by finding more evidence in nature or by building sturdier models. Postponing such backbreaking work could be easy; the very features that add to the institute's appeal—its interdisciplinary and long-term horizons and its ability to embrace scientists from many organizations—mean that few reputations are riding on the institute's success.

John H. Holland, a computer scientist at the University of Michigan and a pillar of the institute, believes a few such efforts are taking root. "These things are rare but extremely important," he says. Adds George A. Cowan, the institute's founder: "Once the questions are sharp, you know you will find the answers." —*Elizabeth Corcoran*

Musseling In

Exotic species hitch rides in ships' ballast water

AQUATIC INTERLOPERS, such as zebra mussels, can cause havoc on arrival. Invasions may be increasing as new trade routes open and ships with larger ballast tanks travel faster. Photo: Renee Stockdale/Animals Animals.

They arrive silently, streaming into harbors, estuaries and coastal waters from the ballast tanks of ships traveling from distant ports. Once here, they may perish—or, with stunning speed, they may alter an entire ecosystem. And recently these exotic species of aquatic organisms are coming more frequently, in greater numbers. They are not welcome.

The zebra mussel, which in the mid-1980s hitched a ride from Europe to Lake St. Clair in the Great Lakes region, is the most infamous of the new invaders. Since then, it has spread through the Hudson, Susquehanna and Mississippi drainage basins, clogging intake pipes. Creatures from ports worldwide are increasingly mixing with their antipodal peers as ballast water sloshes from one ecosystem to another. "It is a far more serious problem than any of us knows," comments Walter R. Courtenay, Jr., a professor of zoology at Florida Atlantic University who studies nonnative fish. "There have been very few ecological studies of the impact of these things."

It is not that Courtenay or his colleagues are xenophobic. The magnitude of the potential biological and economic impact of the multitude of arrivals concerns them. Aquatic mixing on a global scale could lead to a loss of biodiversity, notes James T. Carlton, director of the Williams College-Mystic Seaport Maritime Studies Program, whose bailiwick is ballast water.

Invasions are certainly nothing new. Thousands of species of plants have been either deliberately or accidentally introduced. Interlopers in the animal kingdom include the Asian cockroach and the Norway rat. And for 100 years or so, organisms have been catching ballast-bound rides to and around the U.S. Scores of freshwater, marine and estuarine exotics, as they are called, have been established in that time, Carlton notes.

In the past few years aquatic invasions have received more attention. New trade routes have opened with China, and eastern European ships are now allowed in a greater number of U.S. ports, according to the Transporation Institute in Washington, D.C. The international fleet consists of 39,896 merchant vessels. Ships have become larger, allowing more creatures on board, as well as faster, making the dark passage less taxing.

Since the prolific, pipe-fouling zebra mussel was reported in 1988, two species of spiny-finned fish called gobies from the Black and Caspian seas, an Indo-Pacific crab, an Asian copepod (a tiny crustacean) and a new Black Sea species of zebra mussel have arrived in ballast water and taken up residency. These voyagers are just a few of the hundreds that arrive each month. Carlton has tested ballast water from 159 Japanese ships and found more than 350 living species of plankton alone.

Although the zebra mussel is clearly economically deleterious—an estimated $5 billion will be spent on pest control and equipment repair in this decade—the cost of these other organisms has yet to be assessed. The good news is that the gobies feed on zebra mussels. The bad news is that the second species of zebra mussel may thrive in warmer waters than its predecessor, giving it a much expanded geographic range, explains Edward L. Mills, senior research associate at Cornell University's department of natural resources.

Such uncertainty can be costly. For example, the impact of a 1986 arrival is just being studied. The European river ruffe, a perchlike fish that devours the eggs of other fish, has proliferated rapidly in parts of Lake Superior. "It

is a pest in its native land," says Dennis M. Pratt, a biologist with the Wisconsin department of natural resources. "It will probably be a pest here." According to David Cottingham, director of the ecology and conservation office at the National Oceanic and Atmospheric Administration, $1.2 million has been allocated—on paper—to control the ruffe.

The ruffe and the zebra mussel are just two species that are dramatically altering the Great Lakes. "The lakes have a long history of devastating exotics," Pratt says. For instance, $10.1 million is spent annually to control the sea lamprey, which arrived in the early 1800s and which eats large, commercially important fish. The alewife, which settled several decades after the lamprey, replaced more hardy native herring species. Accordingly, when alewives have major die-offs, their predators—stocked species, such as salmon and trout—die off as well. Stocking the Great Lakes cost $44.7 million in 1988.

There are currently no estimates for the amount of ballast water released nationally, but Carlton says the volume is staggering. In March 1991 alone, for example, he found that 30 million gallons of ballast water were dumped in Norfolk, Va. And ballast water is not the only source of introductions: aquaculture, commercial fisheries and the pet industry are among the other culprits. As a result of escapees from aquariums and fish farms, Everglades National Park "is just turning into a cesspool of exotic fish," Courtenay laments.

Although legislation was passed to protect against invasions such as the zebra mussel, it has been largely ineffectual. A rule in the 1990 Aquatic Nuisance Prevention and Control Act requiring ships to dump ballast before entering the ports is voluntary until this November—and the law applies only to the Great Lakes. In addition, the $2 million appropriated for evaluating the effects of invasions and the feasibility of ballast exchange procedures (in which ships release ballast in the deep sea and take on water there) has not been allocated. "The studies have sort of fallen through the cracks," Cottingham says—although a study of the role of shipping is under way. As for national regulations, "don't expect them anytime soon."

Meanwhile the next invaders may be here. Watch the Chesapeake, notes Carlton, who anticipates the arrival there of more exotics. "Ballast water would not pop up as one of the top 10 concerns for the bay, but it should," he says. "It is just a game of biological roulette we play every day." *—Marguerite Holloway*

Smart Wheels

Driverless vehicles take to the open road

"Take me home!" One day soon Raj Reddy hopes to be able to say that to his automobile. The Carnegie Mellon University robotics researcher is enthralled by the idea of a vehicle that would let someone get behind the wheel drunk or half asleep and still arrive home in one piece.

Reddy is not just another overly enthusiastic automaton builder. A driverless van is already tooling down Pittsburgh's local streets and highways. (A "safety driver" sits behind the wheel in case the automatic pilot malfunctions.) Carnegie Mellon's Navlab prototypes can drive 55 miles per hour—they are programmed not to exceed the speed limit. One can even parallel park once it has found a space. A Mercedes driven by a rival computer in Munich has attained 100 kilometers per hour on the more liberal autobahn.

This performance is a far cry from the early days of autonomous vehicles. In the mid-1980s top speed was under three quarters of a kilometer per hour, and a robotic vehicle was as likely to drive up the side of a tree as down a blacktop path (both are, after all, dark objects with roughly parallel sides tapering toward the top of the robot's field of view).

Advances in computer hardware are responsible for part of the improvement, according to Charles Thorpe, head of the Navlab project. The processing power that cost $750,000 in 1985 is available for less than one hundredth of that price today. More important, however, is that the navigation researchers have learned what information their programs need and can safely ignore the rest.

The road-following module in the Carnegie Mellon vehicle, for example, consists of a neural network that determines whether it should steer right or left by processing a low-resolution video image of the road. Early versions of the network, Thorpe recalls, contained 100 "hidden units"—the parts that sum information from multiple input sources and perform the bulk of the processing. Training sessions ran overnight on a small supercomputer.

Now the network contains a mere four hidden units and performs virtually as well. Furthermore, it runs fast enough so that it can learn as the Navlab is driven along a particular road. The network still needs off-line training, however, to learn how to steer under circumstances that a driver would not voluntarily get into (almost off the edge of the road, for example, or in the middle of a road but at a sharp angle to its direction).

There are still a few serious hitches. Although the road follower operates perfectly at 55 miles per hour, the laser range finder for the obstacle-avoidance module has a strictly limited range. If the van overtakes another vehicle at a relative velocity of more than five miles per hour, it cannot brake fast enough to avoid a collision. So the robotic van always drives in the left lane on the highway, on the assumption that everyone else in that lane will be going faster than it is.

Furthermore, even if a machine can successfully navigate from one place to another, it cannot necessarily do anything when it arrives, comments David P. Miller of the Massachusetts Institute of Technology. When Carnegie Mellon researchers tested the Navlab on a suburban mail-delivery route, for example, they did not even consider trying to open a mailbox and put something in it. R. Peter Bonasso and his colleagues at MITRE in McLean, Va., built a mobile robot for finding and fetching objects that suffers a similar limitation: the robot must rely on a person or another robot to hand it the objects once it has found them.

Miller says that Rocky III, a prototype Mars rover that he and his colleagues developed at the Jet Propulsion Laboratory in Pasadena, Calif., is "the state of the art" in combining navigation with manipulation. It finds its way to a specified area, scoops a gripperful of loose sand from the ground and delivers the "soil sample" to its starting point.

Indeed, Miller suggests researchers should concentrate on specific tasks people want done. Intelligent cruise control might well be one of them, he says, noting that rudimentary systems for collision avoidance are already on the market. A present-day cruise control and navigation system "doesn't steer," he says, "but that's just a reactive hack."

Thorpe prefers problems that pose less of a liability hazard; he predicts the Navlab will be better suited for jobs where "you don't want people around," such as mapping hazardous waste sites. The task requires mobility in rough terrain, but it can be done slowly, and "if you're carrying hundreds of thousands of dollars' worth of sensors, the robot can be expensive, too." Nevertheless, he contends, if people were willing to accept robots as equals, mechanical chauffeurs could be driving regularly by late in the decade. *—Paul Wallich*

Formula for Diabetes?

Cow's milk for infants may contribute to the disease

The National Dairy Board's slogan, "Milk. It does a body good," sounds a little hollow these days. Recently a team of Finnish and Canadian scientists, led by Hans-Michael Dosch of the Hospital for Sick Children in Toronto, found evidence that early exposure to a protein in cow's milk may sometimes lead to juvenile diabetes. A fragment of that protein, it seems, could set up an immunologic cross-reaction that gradually destroys the insulin-making beta islet cells in the pancreas of some susceptible children. Worried parents should not pull the formula bottle out of the baby's mouth yet: nondietary approaches to preventing diabetes, including vaccines, may turn out to be far more practical.

Juvenile diabetes, also known as Type I, or insulin-dependent, diabetes mellitus, has long frustrated researchers trying to understand its genesis. About 25 percent of the population carries at least one genetic marker associated with diabetes, yet less than 1 percent of these people actually becomes diabetic. For that reason, Dosch explains, 85 percent of the newly diagnosed patients come from families that have no history of diabetes.

Beginning in the mid-1980s, some researchers found hints that something in cow's milk might be an environmental trigger for the disease. Epidemiological and clinical data revealed that the incidence of diabetes parallels the per capita consumption of milk in various countries and that breast-fed babies seem less likely to become diabetic.

In other studies, diabetes-prone rats raised for their first month on diets free of cow's milk (a common ingredient in laboratory kibble) did not become ill. Diabetic rats also had high levels of antibodies against bovine serum albumin, the whey protein in milk. In 1991 the Toronto team showed that antibodies against one piece of the albumin molecule cross-reacted with p69, a protein found on beta islet cells.

Dosch's group has now tied the immunologic discoveries in animals closer to humans. As the researchers reported in the *New England Journal of Medicine* this past July, children recently diagnosed as diabetic have blood serum concentrations of antibodies against the bovine serum albumin fragment and p69 that are almost seven times those of their healthy peers. These antibodies vanish within two years after the disease is diagnosed, presumably as the p69 and islet cells disappeared.

It is perfectly normal, Dosch notes, for people to have antibodies against proteins in cow's milk. "Up until the age of about three months, the gut lets through reasonably large protein fragments," he explains. Normally the immune system learns to tolerate foreign proteins that enter the body through the gut—that is, it mounts no more than a mild, ineffective assault against them. In the case of juvenile diabetics, however, Dosch believes this tolerance mechanism has failed for the whey protein fragment.

According to Dosch's hypothesis, that immunologic oversight sets up a tragic chain of events that are triggered by viral infections. When the prediabetic's body releases interferon to mobilize the immune defenses against viruses, that hormone also causes islet cells to stud their surface with p69, the bovine serum albumin fragment look-alike. Primed by exposure to the fragment, the immune system pounces on the islet cells and kills them, too. Each time the body is infected by a virus, more islet cells are lost. "This model really explains for the first time why it takes eight to 10 years for diabetes to develop," Dosch claims.

The case proving bovine serum albumin's role in triggering diabetes is far from ironclad. Although the high levels of antibodies against the fragment show that the immune systems of diabetics are sensitive to the molecule, Dosch and most other diabetes researchers doubt these antibodies destroy the islet cells. Instead *T* lymphocytes, which attack sick or abnormal cells, are the more likely culprits. Dosch and others are now looking for lymphocytes that attack bovine serum albumin.

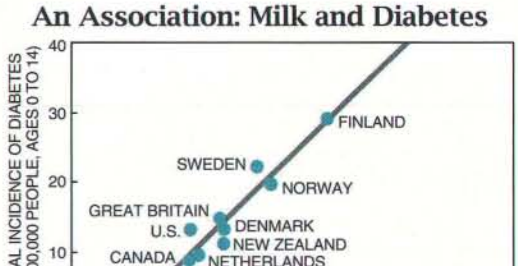
Mark A. Atkinson, a diabetes physiologist at the University of Florida College of Medicine in Gainesville, argues that researchers must also check whether antibodies against the bovine serum albumin fragment are present in people at risk for diabetes. "If they were a cause of diabetes, you would expect to see these antibodies long before the onset of the disease," he says. Atkinson is nonetheless excited by the possibility that Dosch's antibody might be a more accurate diagnostic tool for the disease than any now available.

A definitive test of the cow's-milk theory is scheduled to begin next year, when more than 3,000 newborns with a family history of juvenile diabetes will be put on special diets for their first nine months; the babies will be fed only mother's milk and a dairy formula from which bovine serum albumin has been removed. It will take five to 10 years to determine whether these children have a lower incidence of diabetes, Dosch says. Organizing such a study is difficult, however. One potentially confounding problem is that some evidence suggests dairy products in the diet of the mother can sometimes pass into breast milk. "That makes our trial run rather hard," Dosch says wearily.

Dosch, Atkinson and other authorities all caution strenuously that the work is too premature to mandate a dairy-product ban for infants. "Cow's milk is a good source of protein," Dosch emphasizes. "It's cheap, it's affordable." It is also ubiquitous: bovine serum albumin is an emulsifying and stabilizing agent added to many processed foods. A comprehensive dairy ban could consequently be hard to implement without drastic changes in eating habits in Western countries. "If you were talking about a diet for genetically susceptible people for three months, maybe you could do it," Atkinson ponders. "Bovine serum albumin is in so many different products, it would be difficult if not impossible to keep somebody off that for years."

For these reasons, Dosch and other researchers are looking into other ways to apply their work. One possibility is a vaccine that builds tolerance to bovine serum albumin. "We have a vaccine now that prevents the disease in 80 percent of the rats. Unfortunately, the remaining 20 percent get the disease quicker and worse," Dosch says wryly. An alternative approach would be for genetic engineers to create transgenic cows that have bovine serum albumin not so similar to p69. Changes at only a few amino acid positions would be sufficient, he explains, adding that several dairy and cattle industry groups in the U.S. and Europe have been in touch with him about this possibility. —*John Rennie*

An Association: Milk and Diabetes



SOURCE: Diabetes Care, *1991*

A Matter of Timing

The space telescope takes a few good shots

It may have been coincidental, but the effect seemed quite deliberate. Just as Congress began debating whether to cut funding for big science projects such as the Superconducting Supercollider and Space Station *Freedom,* a flurry of news releases appeared trumpeting the latest discoveries from the *Hubble Space Telescope.*

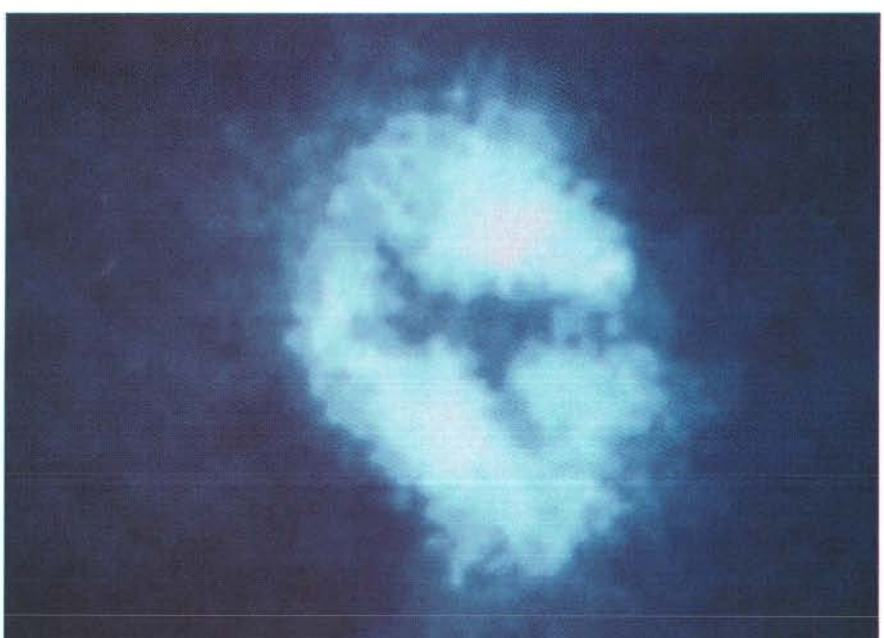
The most recent findings continue the lengthy rehabilitation of *Hubble's* reputation, tarnished by its flawed mirror. "*Hubble* is a great, great telescope," echoes Allan R. Sandage of the Carnegie Observatories, who has been involved in one of the telescope's most far-reaching experiments: an effort to determine the rate at which the universe is expanding.

That rate, known as the Hubble constant, has inspired years of passionate cerebration and argument. Curiously, different groups of astronomers tend to find two distinct values for the Hubble constant, one twice as large as the other. The larger value implies that the universe is about 10 billion years old; the smaller, that the age is closer to 20 billion years. Sandage has long supported the latter number.

Thanks to the space telescope, he now has much more potent ammunition to reinforce his position. In collaboration with Gustav Tammann of the University of Basel in Switzerland and three researchers at the Space Telescope Science Institute, Sandage has used *Hubble* to search for Cepheid variable stars in a faint spiral galaxy known as IC 4182, an oft-studied distance marker. Observations of such stars form the first step in a set of inferences that establish the distance scale, and the rate of expansion, of the universe.

Whereas earth-based observers have never succeeded in monitoring Cepheid variables in IC 4182, *Hubble* enabled Sandage and his colleagues to watch 27 of them. When the researchers examined the results, they deduced a low Hubble constant—exactly as Sandage has maintained all along. "I am very confident in the *Hubble* results. This is the first result they are frightened of," says Sandage, referring to his opponents who favor a younger, more rapidly growing universe. But even he doubts that his latest finding will clinch the debate.

Other space telescope findings also have broad, if less controversial, implications for comprehending how the universe has evolved. Rogier A. Windhorst of Arizona State University and William C. Keel of the University of Alabama, along with several co-workers, turned their attention to an intriguing galaxy known as 53W002. Terrestrial studies of the galaxy's radio emissions suggested that the galaxy might be very young. It is also quite remote, about 10 billion light-years from the earth (assuming a compromise that puts the age of the universe at 15 billion years).



"X" *MARKS THE SPOT where a black hole hides in the core of the nearby Whirlpool Galaxy. Photo: Holland Ford/National Aeronautics and Space Administration.*

Light from exceedingly hot, short-lived stars dominates the optical spectrum of 53W002. Such a large population of evanescent stars implies that the galaxy began to form only about 500 million years ago. On the other hand, *Hubble's* unprecedentedly clear view of 53W002 reveals that the distribution of light within the galaxy closely resembles that of modern, mature elliptical galaxies. Cosmological theories suggest that it takes about 500 million years for a protogalaxy to collapse into such a structure. If so, astronomers are seeing 53W002 at a unique instant during its transition from galactic infancy to adulthood.

Windhorst's observations indicate that despite its young age 53W002 contains very little free, leftover gas, implying that the star-formation process must have been extremely efficient. "Nobody really understands how star formation proceeds," he confesses. "It's really remarkable."

Windhorst also notes that 53W002 must have formed billions of years after the first quasars appeared, confirming other evidence that different galaxies evolve at highly variable rates. Like 53W002, most galaxies probably started to collapse when the universe was already several billion years old, he says. That interpretation would have to be discarded if the universe truly is only 10 billion years old, however, because it would leave too little time for galaxies to have evolved to their present states.

In a related study, Richard Griffiths of the Space Telescope Science Institute is exploring a class of faint but remarkably abundant galaxies that lie a few billion light-years from the earth. The faint galaxies are ubiquitous in all directions, so Griffiths makes use of what would otherwise be wasted observing time, taking wide-field snapshots of random parts of the heavens while one of *Hubble's* other instruments is trained on a specific target.

In the first two images he obtained, Griffiths detected some 2,500 galaxies. He has not yet sorted them out but notes that he sees "a lot of irregular morphology, a lot of interacting systems." Those findings hint that the faint objects may be embryonic, loosely organized stellar systems that subsequently merged to form modern galaxies like the Milky Way.

Hubble has also devoted considerable time to searching for the giant black holes thought to reside in the centers of many galaxies. Views of the galaxies M32 and M87 merely bolstered evidence that other astronomers had already collected from the ground. But a stunning image of M51, more commonly known

as the Whirlpool Galaxy, caught many astronomers by surprise.

Holland Ford of Johns Hopkins University reported the discovery of a peculiar dark "X" that stretches about 100 light-years across the galaxy's bright nucleus. He believes that one stroke of the X is actually an edge-on view of an inclined torus of gas and dust orbiting a massive black hole at the heart of M51. And the other stroke? It could be a second orbiting ring, or it could be matter interacting with energetic jets shooting from the vicinity of the black hole. Ford prefers to think of the other half of it as "a caution sign, showing that we do not yet fully understand what is happening in the center of M51."

Some of *Hubble*'s discoveries hit much closer to home. One recent finding is helping to elucidate the mechanism that powers solar flares, huge plasma eruptions on the sun. For years, scientists have theorized that protons accelerated to high velocities in the solar magnetic field could be instrumental in triggering such eruptions. When those protons crash into the denser, underlying layers, they would pass along their kinetic energy to a flare. The protons should produce a brief, distinctive ultraviolet signal. This signal cannot penetrate the earth's atmosphere, and previous spacecraft could not be pointed at the sun or else could not detect such a short-lived event.

Taking advantage of *Hubble*'s great light-gathering power, Bruce E. Woodgate of the Goddard Space Flight Center, aided by Richard D. Robinson of Computer Sciences Corporation and three other collaborators, finally succeeded in detecting the telltale radiation from downward-streaming protons. He looked not at the sun (which would burn out *Hubble*'s delicate optics) but at AU Microscopium, a red dwarf star in the southern sky. Presumably the mechanisms producing the flares on AU Microscopium are the same as the ones operating on the sun.

Through a stroke of good luck, a large flare erupted on the star during Woodgate's allotted observing time. After sifting through the data, Woodgate's group found a three-second trace of the expected proton signal. Woodgate hopes to coordinate future observations with simultaneous measurements from the ground and from other satellites. The combined results should quantify how much energy the protons carry, thereby demystifying one more aspect of the sun's dynamic and occasionally violent nature.

The current tantalizing batch of *Hubble* findings underscores researchers' bittersweet experience with the telescope. *Hubble* promises even more stunning observations after its optical system is corrected in 1994. If all had gone according to plan, astronomers would be pouncing on those images right now. Instead they have to keep their fingers crossed while *Hubble* weathers the vicissitudes of shuttle schedules and space science funding. *—Corey S. Powell*

The Cosmic Microwave Mirage?

The April announcement that the *Cosmic Background Explorer* satellite (*COBE*) had found slight variations in the microwave background was regarded by many astronomers as long-awaited evidence for the seeds of galaxies in the young universe. Six months later they are still confident that the background and its variations are what remains of the radiation released by the big bang. But controversy has arisen about whether the *COBE* measurements have any relation at all to the structure of the universe billions of years ago.

Lawrence M. Krauss and Martin White of Yale University argue that the variations in the cosmic microwave background may just be a consequence of the theory of relativity: the variations, they say, could be distortions caused by gravitational waves. "One should not jump to the conclusion that what *COBE* is seeing is just density fluctuations," Krauss warns. "At least some or all of it might be gravitational waves."

For decades, astronomers expected to find variations in the faint glow of microwave radiation released as matter coalesced from the cooling universe during its first 300,000 years. Because the present universe is filled with clumps of stars and clusters of galaxies, cosmologists predicted that the early universe should be lumpy, too. And if matter and energy were unevenly distributed in the young universe, the lumpiness should appear as fluctuations in the microwave background. Dense areas of the universe would restrict the emission of radiation [see "The Golden Age of Cosmology," by Corey S. Powell; SCIENTIFIC AMERICAN, July].

Yet the effects of the theory of general relativity may also have influenced the microwave background, as A. A. Starobinsky of the Landau Institute for Theoretical Physics in Moscow first realized in 1979. General relativity describes space as a strong, rigid fabric. But unlike any material, space bends in proportion to the mass it contains. The theory explains how the force of gravity is a consequence of the bending of space. It also predicts that if mass moves rapidly and irregularly, it can produce ripples in the fabric of space. These ripples, known as gravitational waves, compress and expand space as they travel through it.

The likely source of such waves is the so-called inflation of the universe. Within a fraction of a second after the big bang, according to the theory, the universe expanded by a colossal factor of 10^{30}. This rapid inflation could generate gravitational waves in a range of sizes. In the August 10 issue of *Physical Review Letters*, Krauss and White calculate that gravitational waves of sufficient size could account for all the variations in the cosmic microwave background.

In particular, some gravitational waves would be comparable in length to the whole universe. These waves would cause enormous regions of space to expand or compress slightly. Consequently, some regions would emit more radiation than others, and the difference would show up in the microwave background.

The effects, according to Krauss and White, could obscure evidence of the density fluctuations that might be the seeds of galaxies. Other cosmologists do not feel as strongly. "If you make the most likely guess," says Martin J. Rees of the University of Cambridge, "gravitational waves account for less than 10 percent of the variations in the background."

Astronomers may wait for years before they resolve the issue. The answers will have to come from gravitational-wave detectors, particle physics experiments or high-sensitivity measurements of the microwave background. And all three tests may turn out to be projects as ambitious as *COBE* itself. *—Russell Ruthen*

PROFILE: HANS A. BETHE

Illuminator of the Stars

Hans A. Bethe moves slowly now. His left arm dangles at his side, the shoulder wasted more than a decade ago by a degenerative muscle disease. But his husky, low-slung build, that of a chronic mountain climber, still exudes power, gravity. Bethe speaks slowly, too, but with exceptional clarity and precision. His voice, inflected with a German accent even though he came to Cornell University almost 60 years ago, is deep and resonant. Recounting his accomplishments, he seems neither modest nor immodest; he merely describes things as he sees them.

Late in our interview, I can no longer resist asking an embarrassingly trite question. "What do you consider your... your...?" "Greatest achievement?" Bethe says helpfully. I nod. Well, he replies, that ought to be his theory of stellar energy, which he developed in 1938 and which won him the Nobel Prize in 1967. But actually he considers his finest paper to be one he wrote in 1930 on how electrons impart energy to atoms. "It's kind of sad to admit that nothing later was better," he adds.

Before I can stop myself, I ask whether Bethe thinks there is any truth to the old saw that physicists are like athletes, inevitably peaking when they are young. If the 86-year-old Bethe takes offense, he doesn't show it. A faint smile even crinkles his broad, beetle-browed face. "No, not at all," he says. "I think I'm still pretty good." As we speak, in fact, Bethe's secretary is typing a paper in which he presents a new—and, he believes, improved—method for calculating how stars erupt into supernovas.

For a physicist, such longevity is rare indeed, according to John N. Bahcall, an astrophysicist at the Institute for Advanced Study in Princeton, N.J., who has collaborated with Bethe. "I'd be hard-pressed to think of anyone else who contributed significantly to theoretical physics past the age of 50," Bahcall says. Bethe shrugs. "There is nothing as interesting as science," he says. "So as long as the brain lasts, that's what I'm going to do."

If ever a scientist was entitled to rest on his laurels, it is Bethe. Beginning in the 1920s, he wielded quantum mechanics like a laser, illuminating phenomena ranging from elementary particles to dense arrays of nuclei to entire stars. He played a central role in designing the first atomic bomb during World War II, and after the war he helped to nurture nuclear power, a cause he still believes in. "It is more necessary now than ever before because of global warming," he says.

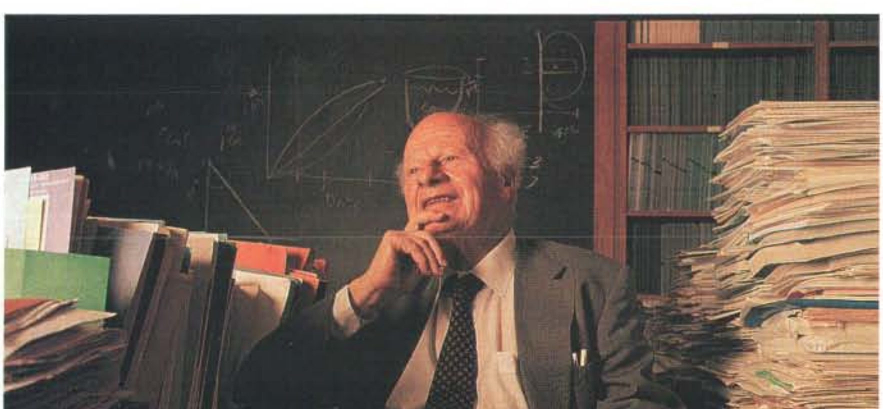
He has also been a persistent advocate of nuclear arms control, helping to persuade the U.S. to ban atmospheric nuclear tests in 1963 and antiballistic missile systems in 1972. Just last year, in an influential essay written for the *New York Review of Books* with Kurt Gottfried, also a physicist at Cornell, and Robert S. McNamara, the former secretary of defense, Bethe urged the U.S. and the Soviet Union to slash their nuclear arsenals to 1,000 warheads each. (Even with scheduled cuts, each side will retain some 5,000 warheads.)

Silvan S. Schweber, a historian at Brandeis University who studied under Bethe and is now writing his biography, notes that Bethe's reputation for integrity as well as brilliance has made him one of the most admired physicists of the 20th century. "I've yet to find somebody, except maybe for Edward Teller, who has something bad to say about Hans Bethe," Schweber adds.

Born and raised in Strasbourg, Bethe studied under some of the founding fathers of quantum mechanics, including Enrico Fermi in Rome and Niels Bohr in Copenhagen, before emigrating to the U.S. in 1935. (Although he was raised in his father's Protestant faith, Bethe had lost a university post in Nazi Germany because his mother was Jewish.) He soon secured his reputation with three review articles that spelled out what was known in nuclear physics and what remained to be known. The papers, which were often referred to as Bethe's Bible, helped to set the agenda for a generation of physicists.

Bethe's Nobel-winning elucidation of starshine came about "rather by accident," he says. In 1938 the Hungarian emigré Teller, who was organizing a conference on astrophysics, invited Bethe to present a paper. "I at first complained, 'Why should I write about astrophysics? There are so many other problems,'" Bethe says. At the time, he explains, he was more interested in the distressing fact that, according to quantum mechanics, the electric charge of electrons should cause them to have infinite energy.

When Teller, who was then one of Bethe's closest friends, persuaded him to attend the conference, he quickly



HANS A. BETHE, who headed the Manhattan Project's theoretical division, later became a forceful proponent of nuclear arms control. Photo: Robert Prochnow.

isolated what he felt was a critical—and solvable—problem in astrophysics. Although other astrophysicists were trying to show how nuclear fusion in stars could give rise to various heavy elements, Bethe's tack was to "forget about that for now and concentrate on how we get the energy." He eventually showed that virtually all of the energy generated by the most brilliant stars stems from a fusion reaction in which hydrogen serves as the fuel and carbon as the catalyst. "That's how I invented the carbon cycle," Bethe says, "or should I say 'discovered.'"

After World War II broke out, Teller and others persuaded Bethe to join in another venture: the construction of a nuclear fission bomb. At the request of J. Robert Oppenheimer, chief scientist of the Manhattan Project, Bethe moved to Los Alamos, N.M., to take over the effort's theoretical division. The group included such prodigies as the late Richard P. Feynman, who was 12 years younger than Bethe.

Recalling this period, Feynman once likened Bethe to a battleship that, escorted by smaller vessels, namely, the younger physicists, was cruising "majestically" into the unknown. The admiration was mutual: Bethe says the Hungarian mathematician John von Neumann was probably the most brilliant physicist he has known, and Bohr was the greatest, in terms of actual accomplishments, but "Dick Feynman combined brilliance with greatness."

Bethe is reluctant to describe his role in quelling fears that the first atomic explosion might ignite a global conflagration; the incident's significance has been greatly exaggerated since it was first publicized in 1975, he says. But he briefly recounts what happened. In 1942 Teller had warned that the heat of a nuclear blast might trigger runaway fusion of atmospheric nitrogen and set the entire earth ablaze. Oppenheimer and others took the suggestion quite seriously, Bethe notes. But he "sat down and looked at the problem and found that it was just incredibly unlikely."

Later, a more detailed analysis by two other physicists at Los Alamos corroborated Bethe's conclusion and "put the question at rest," he says. He recalls that just before Trinity, the first test of the atomic bomb, Fermi jokingly asked if anyone wanted to bet whether the explosion would set the world on fire. Bethe himself was sure that such an event was not possible. Didn't he worry at all, during the countdown, that a cataclysm might ensue? "No," he replies. His only concern was that the initiator, which he had designed, would not work. It worked.

Bethe says that as World War II gave way to the cold war, he opposed the development of the vastly more powerful fusion bomb, also called the hydrogen bomb or simply "the super." Although he believed fission-based weapons were needed to counter Soviet aggression, he felt that the U.S., by renouncing the super, might have convinced the Soviet Union to do so as well. Now he is not so sure, since the Russian physicist Andrei Sakharov revealed in his memoirs that the Soviets "were determined to build one no matter what we did."

At any rate, Bethe's ambivalence toward the super strained his relationship with Teller, the bomb's most enthusiastic advocate. The two physicists became still more estranged in 1954, after Teller testified before a hearing board set up by the Atomic Energy Commission to investigate whether Oppenheimer—who shared Bethe's misgivings about the super—was a security risk. Teller "gave a very convoluted testimony," Bethe says, "but it was clear that he stood against Oppenheimer."

Three years ago Bethe, then 83, decided to learn lattice gauge theory, one of the most challenging disciplines in physics.

During the last decade, Bethe and Teller have clashed repeatedly over the Strategic Defense Initiative, also called Star Wars, Teller's plan to create a shield against nuclear missiles. After President Ronald Reagan launched the initiative in 1983, Bethe criticized it as technically impossible and politically destabilizing. Can the two former friends find anything to agree on any more? "We had a very nice talk together" at a conference two years ago, Bethe says, "mostly concerning children."

Bethe's involvement with political issues never kept him from pursuing his first love, pure physics. In the late 1940s he suggested a way to eliminate the impossible predictions of infinite energy arising from quantum mechanical descriptions of electrons. Building on Bethe's ideas, various younger physicists, notably Feynman, who followed Bethe from Los Alamos to Cornell after World War II, developed quantum electrodynamics, an astonishingly successful theory of all electromagnetic effects.

In the 1950s and 1960s Bethe devised potent techniques for calculating the properties of dense agglomerations of neutrons and protons, also known as nuclear matter. The topic still fascinates him. In fact, three years ago he "apprenticed" himself to another physicist, Gerald E. Brown of the State University of New York at Stony Brook, in order to learn lattice gauge theory. The theory, which predicts how nuclear matter is transformed at extremely high temperatures into a plasma of quarks and gluons, is one of the most complicated and challenging in physics. "I'm interested in learning new things," Bethe says.

Since his retirement from teaching duties in 1975, however, Bethe has devoted most of his energy to the subject he once spurned: stars. In the past few years, he has written three papers with Bahcall of the Institute for Advanced Study trying to explain why the sun spews out fewer neutrinos than predicted by standard physics. Bahcall and Bethe have championed the Mikheyev-Smirnov-Wolfenstein theory (named after its originators), which holds that certain solar neutrinos are transformed before they reach the earth into variants that are more difficult to detect.

Bethe has also attempted to understand how stars explode. It is a little-known embarrassment in astrophysics that—in spite of the enormous advances in observation, theory and computer technology—the latest computer models of supernovas either yield puny half-explosions or complete duds. Bethe has struggled with the problem for 14 years, but now he thinks he might have solved it.

His proposal calls for simplifying current supernova models in at least two ways. Rather than calculate neutrino production from the bottom up, he plugs in a number inferred from observations of the giant supernova of 1987. He also employs a more efficient method of calculating the net energy generated in a supernova by two competing nuclear reactions involving neutrinos and electrons, respectively. These simplifications yield a model that explodes on cue; it is also so streamlined that its results can be calculated without a computer. "I can do everything with algebra," Bethe remarks. "So that makes me very happy."

Although we have been talking for four straight hours now, Bethe is having a hard time hiding his excitement over the proposal. He emphasizes that it has not yet been reviewed; indeed, perhaps at this very moment, his secretary is preparing to send out preprints to a select group of experts. But he is confident that it will fare well. It's "a pretty good paper," he says—almost as good, even, as his all-time favorite, the one he wrote back in 1930. —*John Horgan*

SCIENTIFIC AMERICAN

How Many Species Inhabit the Earth?

The sad truth is that no one knows. The answer is relevant to efforts to conserve biological diversity and could illuminate crucial questions about evolution and management of the environment

by Robert M. May

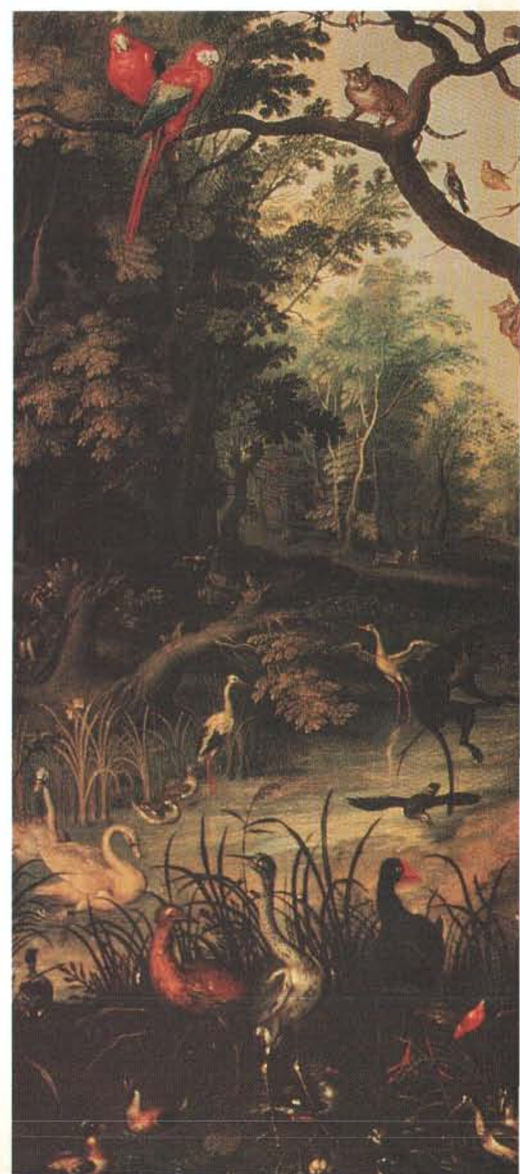
If an extraterrestrial explorer were to land on the earth, what is the first question it would ask? The alien would, I think, inquire about the number and variety of living organisms on this planet. Given that the earth's physical attributes derive from universal and essentially deterministic laws, the presumably well-traveled visitor would probably have seen countless similar worlds throughout the universe. But the warp of evolutionary forces and the weft of chance that crafted the rich tapestry of life on the earth are almost certainly unique.

Surprisingly, we humans cannot even approximately answer the alien's query. Despite more than 250 years of systematic research, estimates of the total number of plant, animal and other species vary widely, all the way from three million to 30 million or more. Because no central archives exist, no one even knows how many species have already been named and recorded.

ROBERT M. MAY has been Royal Society Research Professor at the University of Oxford and at Imperial College, London, since 1989. He was trained as a theoretical physicist at Sydney University. During the early 1970s, May became interested in the dynamics of animal populations and in the relations between stability and complexity in natural communities. In 1977 he moved to Princeton University, where he served as a professor of zoology and as vice president for research. May is currently studying the factors influencing the diversity and abundance of species, in particular evolutionary interactions between parasites and their hosts.

I find the current state of affairs particularly distressing in light of the rapid rate at which wild habitats are being destroyed. A knowledge of the total number and distribution of species is fundamental for developing a rational program to conserve as much as possible of the remaining biological diversity. The British government expressed the argument with admirable clarity in a 1990 white paper: "The starting point for this Government is the ethical imperative of stewardship which must underlie all environmental policies.... We have a moral duty to look after our planet and to hand it on in good order to future generations."

Political leaders face a baffling assortment of environmental concerns that cry out for a better understanding of how ecosystems change as constituent species die out and are succeeded



***THE TERRESTRIAL PARADISE,* painted by Jan Brueghel the Elder, concentrates on the species most useful or attractive to humans; birds and mammals are therefore grossly overrepresented. Such historical prejudices are reflected in the fact that modern tallies of bird and mammal species are nearly complete, but the total diversity of organisms such as insects, spiders, fungi, nematodes and bacteria remains poorly known.**

by other forms of life. Such an inquiry explores the fundamental relation between diversity and the stability of a biological community. It also touches on other compelling issues, such as predicting changes in the earth's climate. After all, the earth's oxygen-rich atmosphere was originally produced by living organisms, a fact that underlines the extent to which ecosystems and the atmosphere are intertwined.

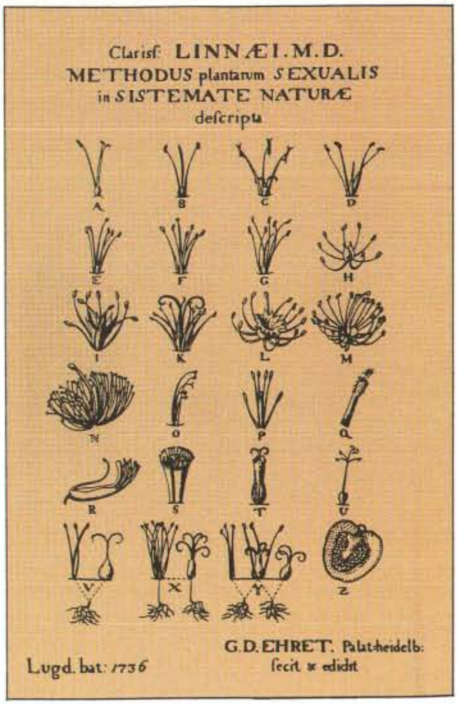
More immediately, utilitarian reasons for counting and cataloguing species are also noteworthy. A considerable fraction of modern medicines has been developed from biological compounds found in plants. Society would be well advised to keep looking at other shelves in the larder rather than destroying them. Many nutritious fruits and root crops remain largely unexploited; cultivating them could expand and improve the global food supply.

Even within familiar genera of crop plants, researchers continue to uncover new geographic varieties. Such variants contain the raw material from which more productive and disease-resistant strains can be created by selective interbreeding or by genetic engineering. The triumphs of modern intensive agriculture have been accompanied by a dangerous narrowing of the diversity of the plants that farmers use, which increases the susceptibility of crops to disease and climatic variability. The likelihood of shifts in the global environment gives fresh emphasis to the desirability of conserving the gene pool and to exploring the use of different plants.

Efforts to view the organic world as an ordered system date back at least to Aristotle. The task of systematically naming and recording species, known as taxonomy, was begun by the great natural scientist Carolus Linnaeus, in Uppsala, Sweden. The canonical 10th edition of his book *Systema Naturae,* which recorded some 9,000 species of plants and animals, was published in 1758—a full century after Isaac Newton had arrived at an analytic and predictive understanding of the laws of gravity, based on centuries of detailed astronomical observations.

Since then, taxonomists have added species to Linnaeus's list at highly disparate rates in different biological categories. By far the most attention has been lavished on animals endowed with the charm of feathers or fur. Scientists have nearly completed cataloguing such species. For example, less than a century after Linnaeus's work, half of the 9,000 known species of birds had been recorded. Researchers currently uncover only about three to five new bird species a year. A similar situation holds for the 4,000 or so species of mammals, although on average about 20 species and one genus are discovered annually. About half of those are truly undiscovered species (mostly rodents, bats or shrews), whilst the others result from reclassifications of old species based on updated biochemical findings.

Quite a different pattern emerges for organisms other than birds and mammals. A record of the growing number of known species of arachnids and crustaceans (essentially, arthropods other than insects) shows that relatively high rates of discovery prevailed in Victori-



CAROLUS LINNAEUS (*left*) established in the mid-18th century the nomenclature and classification system of modern taxonomy. An illustration from his book *Systema Naturae* (*right*) shows how he organized plants based on leaf shapes.

an times, followed by a prolonged lull. More than half of the current total has been added within just the past few decades. In a recent study, Peter M. Hammond of the Natural History Museum in London demonstrated that from 1978 to 1987 the number of known species of birds increased an average of just 0.05 percent a year. Over the same period, the recorded number of species of insects, arachnids, fungi and nematodes expanded by 0.8, 1.8, 2.4 and 2.4 percent a year, respectively.

The varying rates of discovery to some extent reflect different sizes in the work force of taxonomists dedicated to studying each group of organisms. Accurate statistics on the work force are hard to obtain, but a cursory survey of researchers in Australia, the U.S. and Britain by Kevin J. Gaston of the Natural History Museum and me offers some perspective. We found that if N represents the average number of taxonomists studying each species of tetrapod (all vertebrates other than fish), then there are approximately $0.3N$ taxonomists for each species of fish and only from $0.02N$ to $0.04N$ for each invertebrate. Some 10,000 taxonomists work in North America; the global total is perhaps three times greater.

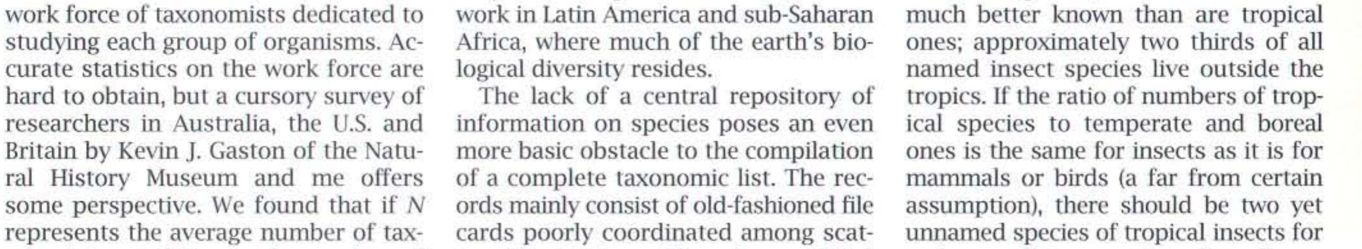
Overall, there are twice as many taxonomists for each recorded plant species as for each animal. Within the animal kingdom, the average vertebrate species receives about one order of magnitude, or factor of 10, more taxonomic attention than does the average plant, whereas the average invertebrate receives about an order of magnitude less. The distribution of taxonomists is manifestly ill matched to the species richness of the various taxa. Moreover, only about 4 percent of taxonomists work in Latin America and sub-Saharan Africa, where much of the earth's biological diversity resides.

The lack of a central repository of information on species poses an even more basic obstacle to the compilation of a complete taxonomic list. The records mainly consist of old-fashioned file cards poorly coordinated among scattered institutions. There is no official count of the species that have already been named. Scientists know far more about (and spend vastly more money studying) the systematics of stars than the systematics of earthly organisms. Consequently, they have as good a knowledge of the number of atoms in the universe—an unimaginable abstraction—as they do of the number of species of plants and animals.

According to the best estimates, taxonomists have identified 1.5 to 1.8 million species, but the list is obviously far from complete. An assortment of approaches—some empirical, others theoretical—has been used to determine, at least approximately, the real total. Even the lowest estimates indicate the existence of three million species, so many that present methods will not suffice to discover and catalogue them all within a reasonable time.

Several authors have estimated the global species total by extrapolating from trends in recording species. Such work is prone to divergent interpretations depending on one's statistical procedures. A recent study, conducted using separate statistical projections from trends of discovery for each major biological group, concluded that there are six to seven million species. Another approach, which drew on the opinions of experts in each group, implied the existence of upward of five million species.

Many projections rely on a simple intuitive argument, based on the relative abundances of species in different taxonomic categories. For well-studied groups such as birds and mammals, tropical species are about twice as common as temperate or boreal ones. But among insects, which account for the majority of all recorded species, northern faunas are much better known than are tropical ones; approximately two thirds of all named insect species live outside the tropics. If the ratio of numbers of tropical species to temperate and boreal ones is the same for insects as it is for mammals or birds (a far from certain assumption), there should be two yet unnamed species of tropical insects for each named temperate or boreal species. Applying the above logic expands



RICHNESS OF LIFE ON EARTH takes on an unfamiliar appearance when organisms are ranked according to their contribution to the total biodiversity, as measured by the number of species in different groups. Insects account for more than one

the 1.5 to 1.8 million recorded species to from three to five million.

A more direct way of estimating the global species total—especially the number of tropical insect species—involves thoroughly sampling the organisms living in some region that has been relatively unstudied and then determining what fraction of those plants and animals has already been described. That method presents quite a challenge: even within a limited area it is difficult to sample all of the tropical insects. Identifying and classifying them presents an even more onerous task. And one must worry whether the site or group is typical of the general patterns of species distribution.

Ian D. Hodkinson and David S. Casson of the Liverpool Polytechnic in England assessed the hemipterans, or true bugs, in a moderately large, topographically diverse region of tropical rain forest in Sulawesi, Indonesia. They found a total of 1,690 species of terrestrial bugs, 63 percent of which were previously unknown. If this fraction is representative of insects generally, then one could infer from the 900,000 recorded species of insects an actual total of two to three million species.

Hammond carried out a different version of this local-population technique. He observed that the 22,000 insect species found among the well-inventoried fauna of Britain (where generations of clergymen have given more attention to their six-legged flocks than to their two-legged ones) include 67 species of butterflies. Many naturalists accord butterflies the honorary status of birds, and hence the 17,500 known species represent a nearly complete inventory. The true number almost surely is no more than 20,000.

If the taxonomic composition of the insect fauna of Britain resembles the global composition, the world should contain a total of about six million (22,000 times 20,000, divided by 67) insect species. Such scaling-up estimates suffer from a built-in uncertainty. One cannot be certain that any individual group of insects or any particular place is typical of the insect population in general.

Terry L. Erwin of the Smithsonian Institution directed a particularly intriguing study in which he and his co-workers scrutinized beetle faunas in the canopies of tropical trees. Beetles are distinguished from other insects by their hard, winglike plates that cover functional inner wings. Approximately one in five of all recorded species are beetles, a fact that prompted the British geneticist J.B.S. Haldane to joke that his studies of the natural world had taught him that the Creator has "an inordinate fondness for beetles." Furthermore, tropical forests are rich sites of biodiversity. Although they cover only about one sixteenth of the earth's land area, tropical forests may harbor at least as many species as the rest of the world's regions added together. Even so, few researchers anticipated the stunning implications of Erwin's work.

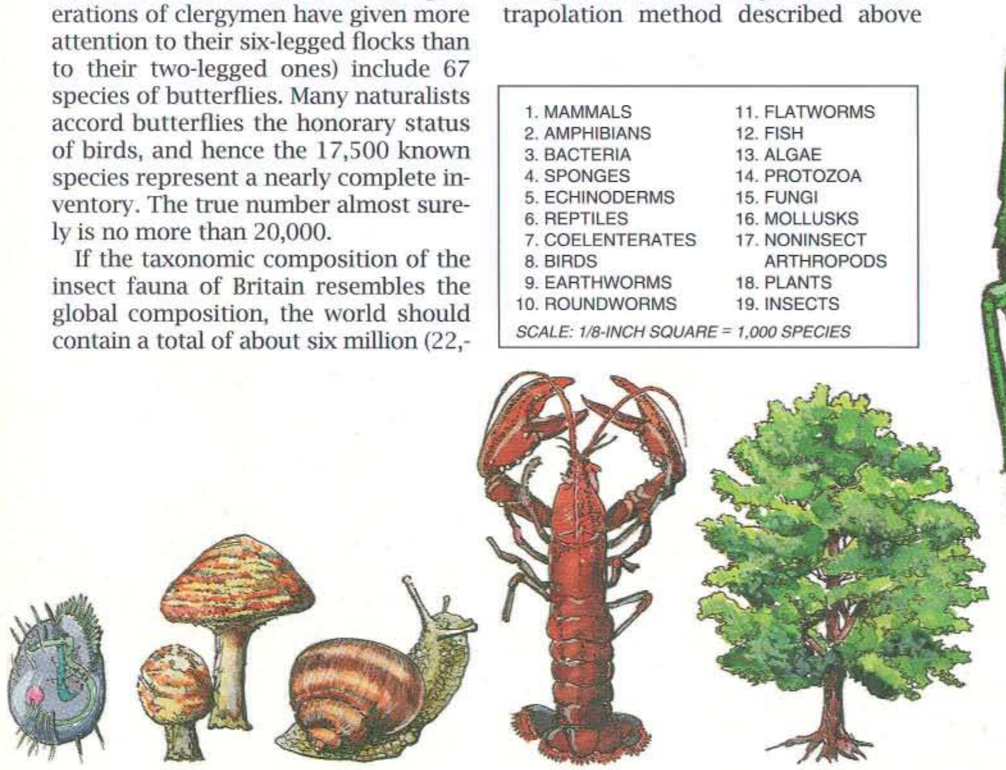
Using an insecticidal fog, Erwin collected canopy beetles in Panama from *Luehea seemannii* trees, tropical cousins of the linden. Over three seasons, Erwin found a remarkable total of some 1,200 beetle species. He has not yet sorted them out to determine how many of the species are uncatalogued, so the extrapolation method described above cannot be applied. Instead Erwin used the following chain of argument.

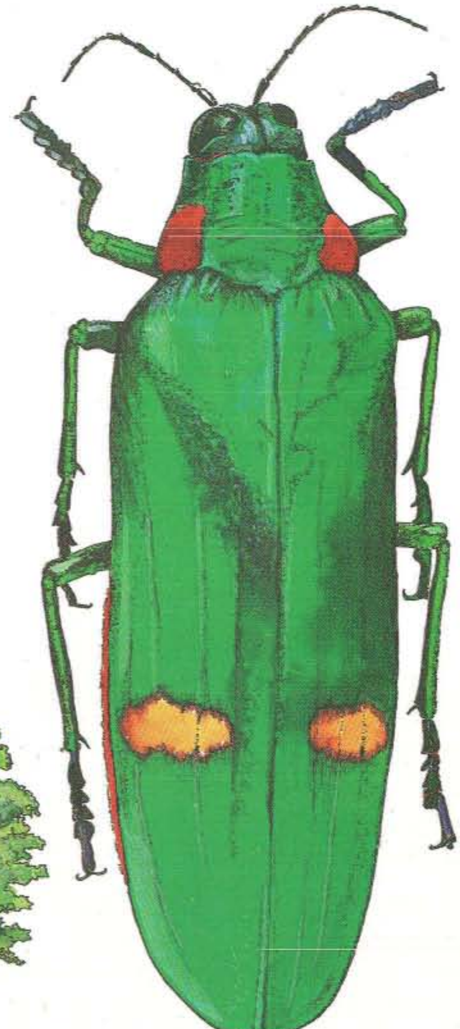
First, he needed to know how many of the beetle species he collected live specifically on *L. seemannii*, as opposed to being distributed across many kinds of trees. Erwin guessed that around 20 percent of the herbivorous beetles (the largest group in his sample) are specialized to each tropical tree species. On that basis, he estimated that one kind of tree holds an average of 160 species of canopy beetle. Second, Erwin inferred the plenitude of all insect species from the species density of canopy beetles. Forty percent of known insects are beetles; if this proportion applies in tropical tree canopies, then 400 kinds of canopy insects occupy each tree species.

Third, Erwin supposed that the canopy contains two thirds of the insect species on the tree, implying a total of 600 insect species on every variety of tropical tree. Finally, he cited a widely

1. MAMMALS	11. FLATWORMS
2. AMPHIBIANS	12. FISH
3. BACTERIA	13. ALGAE
4. SPONGES	14. PROTOZOA
5. ECHINODERMS	15. FUNGI
6. REPTILES	16. MOLLUSKS
7. COELENTERATES	17. NONINSECT ARTHROPODS
8. BIRDS	18. PLANTS
9. EARTHWORMS	19. INSECTS
10. ROUNDWORMS	

SCALE: 1/8-INCH SQUARE = 1,000 SPECIES





half of all known species, even though they certainly have not all been catalogued. In contrast, there are only 4,000 species of mammals, about one quarter of 1 percent of the roughly 1.5 million recorded species.

accepted estimate that the earth supports 50,000 species of tropical trees. Multiplying 600 insect species times 50,000 tree species yields 30 million kinds of insects. The number of all species worldwide would obviously have to be much larger still.

Each link in Erwin's chain of logic involves extreme uncertainty. In my opinion, tropical beetles as a group are likely to be significantly less specialized than are temperate ones, in which case Erwin's figure of 20 percent should be closer to 2 or 3 percent. On the other hand, Erwin probably underestimated the species density on the noncanopy part of the tree, which I suspect may hold as much as two thirds of the insect species.

If one incorporates what I consider to be more realistic numbers, the calculations imply the existence of only three to six million insect species. I nevertheless believe that Erwin's work is important, both in providing a focused approach to the problem of estimating how many species exist and in emphasizing how taxonomic questions interweave with ecological ones. Erwin's efforts do not settle the question of how many species exist so much as they define an agenda for research.

The ultimate aim in recording biological diversity is to build a factual foundation for answering basic questions about evolution and ecology. Taxonomic lists serve as points of departure for studying the structure of food chains, the relative abundance of species, the number of species or total number of organisms of various physical sizes, and general trends in the numbers and distribution of living things. Some of those studies suggest broad rules that make it possible to estimate the number of species in other, independent ways.

One such technique derives from observing patterns of how many species of land-dwelling animals fall into various body-size categories. In general, taxonomists find that for each 10-fold reduction in length the number of species fitting into that size category increases 100-fold. The pattern holds for animals ranging in length from a few meters down to around one centimeter.

Below one centimeter the relation begins to fall apart, possibly because of the highly incomplete record of such tiny terrestrial animals. If the size-versus-species density pattern were arbitrarily extrapolated down to animals having a characteristic length of around one millimeter—an arbitrary dividing line between macroscopic and microscopic life—it would imply a global total of some 10 million species of land-dwelling animals. Such a purely phenomenological assessment would be more persuasive if taxonomists confidently understood the physiological, ecological and evolutionary factors responsible for body-size distributions.

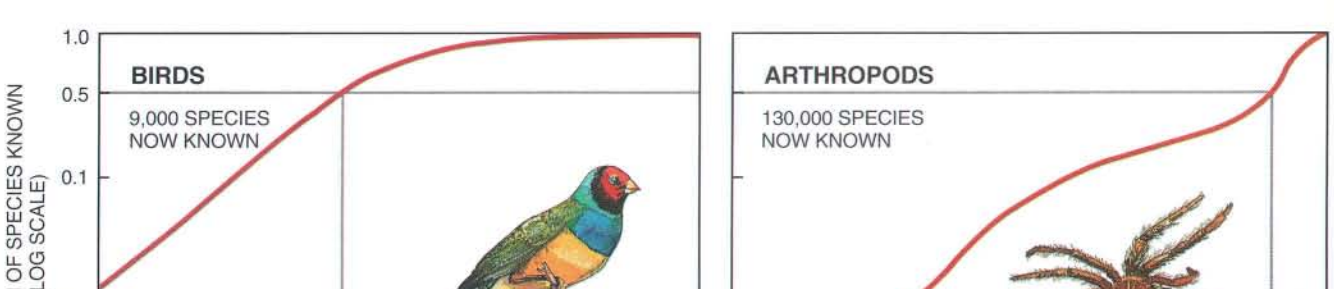
(The size distribution rule sheds light on Noah's problem with the size of the Ark. Pundits have often suggested that Noah would have faced impossible obstacles getting all the myriad insect species on board. But if a 10-fold decrease in size—equivalent to a 1,000-fold decrease in volume—results in only 100 times as many species, then the difficulty really lay in getting the largest creatures on board.)

General patterns in the structure of the food chain provide the basis for another kind of species tally. Photosynthetic plants create the raw organic material that forms the first link in the food chain. If one could obtain detailed statistics on how many other life-forms each kind of plant can support, then the total number of species could be derived from the comparatively complete counts of plant species.

Although scientists remain a long way from this grail, Gaston has extended earlier research on food chains by collecting available evidence on the average numbers of insect species associated with each plant species in communities of highly disparate sizes and locations. On average, he finds approximately 10 insect species for each plant species. Given the fairly secure estimates of about 270,000 species of vascular plants overall, Gaston crudely derives a total of three million insect species.

The discussion up to this point has dwelt mainly on calculating the numbers of terrestrial insect species, and for good reason: those creatures already constitute more than half of all recorded species, and yet they are far from being completely catalogued. Several other taxonomic groups may rival the diversity of the insects, however. In particular, organisms that are small or unglamorous, or both, have probably been disproportionately overlooked.

David L. Hawksworth of the International Mycological Institute in Kew, England, has proposed an upward revision of the total number of fungal species at least as striking as the one suggested by Erwin's work on insects. Hawksworth begins by noting that taxonomists have recorded some 69,000 species of fungi. In Britain and other well-scrutinized northern European regions, fungal species outnumber vascular plant species by about six to one. If that ratio holds true around the globe, then the 270,-



NUMBER OF SPECIES recorded by taxonomists has grown at highly divergent rates for different forms of life. Half of all known bird species had been discovered by 1845; only a few species are added to the list each year (*left*). For arachnids and crustaceans, in contrast, most of the recorded species have been found just since the 1960s (*right*).

000 catalogued vascular plants coexist with 1.6 million species of fungi, more than 20 times the number now known.

Biological patterns seen in temperate zones might not apply to tropical communities, of course. Tropical fungal species may each associate with a broader range of plants than do those in the temperate zones; such a discrepancy would lower the ratio of fungal to plant species. Conversely, Hawksworth ignores fungi associated with insects rather than with plants, which steers him toward a low estimate. In recent studies of particular tropical sites, the proportion of previously undescribed species of fungi hovers around 15 to 30 percent, significantly below the 95 percent one might expect if Hawksworth's calculation were correct. On the other hand, the studies are not even close to being synoptic, so it may be unfair to expect them to uncover the full panoply of uncatalogued species.

The sketchy knowledge of the number of fungal species belies the fact that they are vital elements of most ecosystems, aiding the decomposition of organic material and the formation of new soil. Fungi have undoubtedly shaped the development of biological diversity, first by helping plants colonize dry land and then, especially through symbiotic relations, by assisting in the spread and diversification of vascular plants, insects and other organisms. Such an important part of the earth's biota deserves more attention.

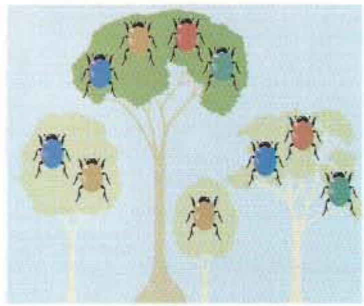
Nematodes are probably the least well inventoried of all animals visible to the naked eye. These tiny worms live as parasites in plants and animals and as free-living creatures in freshwater and marine settings. As of 1860, only 80 species of nematodes had been recognized. Today's total stands at about 15,000. Recent investigations of nematodes on land and in fresh water suggest that recorded species account for only a tiny fraction of the total population. Other studies indicate even greater diversity in marine environments. Few taxonomists would quarrel with Hammond's assessment that the number of nematode species is at least some hundreds of thousands.

The smallest living creatures—those invisible to the unaided eye—probably also contribute greatly to the global cache of species. Microorganisms such as protozoa, bacteria and viruses account for only about 5 percent of recorded living species. Recent work has revealed, however, that the diversity among natural populations of microbial organisms is far greater than that seen in studies of familiar laboratory cultures. One study of RNA from a mat of photosynthetic bacteria in a hot spring at Yellowstone National Park found eight distinct genetic sequence types, none of which matched the 12 strains of laboratory-cultured bacteria that had been believed to be characteristic of such mats. Only one of the eight sequences bears any close resemblance to those belonging to a recognized bacterial phylum.

Biologists who have examined ribosomal RNA sequences in natural marine microbe populations have arrived at broadly similar results. Those studies are, in some ways, even more remarkable than the revelations about tropical canopy faunas because they demonstrate how astonishingly little taxonomists know about the simplest, most common forms of life.

The task of categorizing bacteria and viruses is complicated by the fact that different strains readily exchange genetic material and that a single parent can clone itself to create an entire population. Also, some viruses mutate noticeably from year to year. Basic notions about what constitutes a species are therefore more vague for microbes

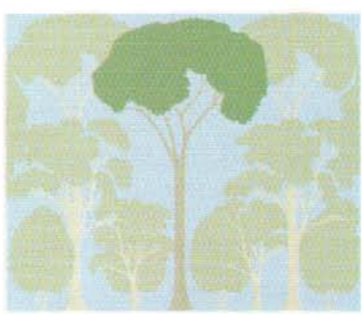
Bootstrap Method of Estimating Insect Diversity



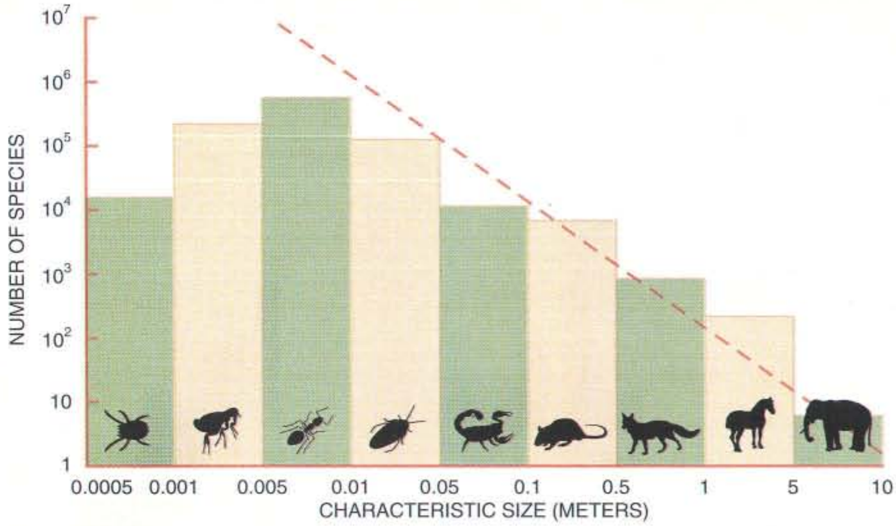
The number of beetle species found in the canopies of a particular tree species forms the basis of the estimate. Strictly speaking, the number of beetle species specialized to those trees is defined as all the beetle species living only on that kind of tree, half of the species on two kinds of trees, one third of the species on three kinds of trees and so on.

Nonbeetle insects account for about 60 percent of all insect species worldwide. Terry L. Erwin of the Smithsonian Institution roughly estimated that 160 beetle species are specialized to the canopies of one kind of tropical linden tree. If beetles constitute 40 percent of the insect diversity, then there must be 400 species of insects specialized to those trees.

Other parts of the tree, including the trunk, roots and surrounding leafy soil, also harbor insects. Erwin guessed that two thirds of all insect species live in the canopy, which would imply the existence of 600 insect species for each tree species.



Tropical forests contain about 50,000 species of trees. If, on average, 600 insect species are associated with each kind of tree, then the tropical forests contain an astounding 30 million (50,000 times 600) insect species. Erwin's result is tentative, however, and may be too high by a factor of five or more.



SIZE RELATION offers a crude way to guess the number of species. In general, there are systematically more species of small animals than of large ones. The relation appears to break down for creatures less than one centimeter long, perhaps because taxonomists tend to overlook such species. If the relation holds for organisms as small as a millimeter long, there are about 10 million species.

than for vertebrates. Manfred Eigen of the Max Planck Institute for Biophysical Chemistry in Göttingen and Peter Schuster of the University of Vienna argue that the basic unit of classification of many viruses should be the quasispecies, a well-defined set of distinctive RNA sequences. Natural selection acts not on viral species as such but on the quasispecies swarm.

The contribution of microorganisms and nematodes to the overall gene pool may be surprisingly large. One rather flippant, but not entirely unreasonable, view is that every species of arthropod and vascular plant (which together account for the vast majority of all recorded species) has at least one parasitic nematode, one protozoan, one bacterium and one virus specialized to it. If so, the previously described species counts should automatically be scaled up by a factor of five. In that case, the total could potentially exceed 100 million, although I doubt the number is so large.

Species counts are useful because the species is a fairly concrete and readily defined measure of genetic uniqueness. But many evolutionary questions or practical purposes direct a researcher's interest toward diversity at higher or lower taxonomic levels.

Moving up the taxonomic hierarchy, from species through genera, families, orders and classes to phyla, emphasizes ever more fundamental genetic variations. For instance, less than 15 percent of all recorded species dwells in the ocean. But more than 90 percent of all classes of organisms and essentially all phyla are represented in the sea. Indeed, two thirds of all phyla is found only in the sea. Hence, in terms of the range of body plans of organisms, the ocean is much more diverse than the land.

At the other end of the hierarchy, conservationists are often interested in the genetic differences among individuals within a single species. When a species nears extinction, much of its genetic diversity disappears, a loss that cannot be reversed by breeding programs. Investigations of diversity within a species therefore offer a measure of the degree to which endangered species have already been irretrievably depleted. Genetic studies can also suggest whether a species may have approached extinction in the not too distant past.

Improving the catalogue of life will require a huge coordinated and sustained effort. I believe such dedication is warranted. Conservation will increasingly necessitate intervening in and managing ecosystems, along with making agonizing choices of where to concentrate such efforts. Those actions will demand better information than is currently available.

The human race also needs to understand the diversity of living things—how much there is and why—for the same reasons that compel people to grope for an understanding of the origin and fate of the universe or of the sequence of molecules in the human genome that code for our own self-assembly. Unlike those other quests, the task of studying and conserving biological diversity has a strict time limit. Each year from 1 to 2 percent of the world's tropical forests are destroyed to create farmland and provide fuel and raw material. At that rate, the tropical forests will be gone in 50 years or so.

The time constraints mean that taxonomists must develop imaginative approaches for cataloguing diversity. One method is to enlist the help of nonprofessionals (so-called barefoot taxonomists) to collect and classify specimens using rough-and-ready techniques. They can gather most of the wanted information in a fraction of the time that would be needed to satisfy the canons of traditional scholarship. Simplified taxonomic research programs are being pioneered in Costa Rica by Rodrigo Gamez and others at the National Institute of Biodiversity; Australia is sponsoring similar work as part of an effort to codify the national collection of species.

Other approaches make use of high technology. Researchers could collect, store and distribute species information on computer disks; efficient programs could be developed to cross-check a species against an organized data base, thereby revolutionizing the time-consuming task of separating and classifying new species. Reference images could be made into holograms that would be readily available. The centralized Australian data base already incorporates some of those features. If widely and creatively applied, modern information technology could make extensive European collections—the legacy of empire—accessible to underequipped researchers in the tropics.

Future generations will, I believe, find it incomprehensible that in the late 20th century, Linnaeus continued to lag far behind Newton. They may also be saddened—as people should be today—that our society has devoted so little money and effort toward quantifying and conserving the forms of life that define the earth's unique glory.

FURTHER READING

HOW MANY SPECIES? Robert M. May in *Philosophical Transactions of the Royal Society of London,* Vol. 330, Series B, pages 293–304; November 29, 1990.

INSECT ABUNDANCE AND DIVERSITY IN THE DUMOGA-BONE NATIONAL PARK, N. SULAWESI. Peter M. Hammond in *Insects and the Rain Forests of South East Asia (Wallacea).* Edited by W. J. Knight and J. D. Holloway. Royal Entomological Society of London, 1990.

FROM GENES TO ECOSYSTEMS: A RESEARCH AGENDA FOR BIODIVERSITY. Report of an IUBS-SCOPE-UNESCO Workshop. Edited by Otto T. Solbrig. Paris, International Union of Biological Scientists, 1991.

TAXONOMY OF TAXONOMISTS. Kevin J. Gaston and Robert M. May in *Nature,* Vol. 356, No. 6367, pages 281–282; March 26, 1992.

Swissair Customer Portrait 161: André Roth, photographer, Zurich, photographed by Marc Schmid.

As a photographer, André Roth is used to registering everything at a glance. And he sets standards that are far above the average. His assignments take him to all parts of the world. That's how he's come to know – and appreciate – Swissair's MD-11. He reckons it's great to have only one seat next to him in the spacious Business Class. Now he really has room to stretch out.

Quantum Cryptography

For ages, mathematicians have searched for a system that would allow two people to exchange messages in absolute secrecy. Quantum mechanics has now joined forces with cryptology to achieve a major step in that direction

by Charles H. Bennett, Gilles Brassard and Artur K. Ekert

In his classic short story "The Gold Bug," published in 1843, Edgar Allan Poe explains the rudiments of code breaking and ventures the opinion that the human mind can break any cipher that human ingenuity could devise. During the subsequent century and a half, the contest between code makers and code breakers has undergone reversals and complications that would have delighted Poe. An unbreakable cipher was invented in 1918, although its unbreakability was not proved until the 1940s. This cipher was rather impractical because it required the sender and receiver to agree beforehand on a key—a large stockpile of secret random digits, some of which were used up each time a secret message was transmitted. More practical ciphers with short, reusable keys, or no secret key at all, were developed in the 1970s, but to this day they remain in a mathematical limbo, having neither been broken nor proved secure.

A recent unexpected development is the use of quantum mechanics to perform cryptographic feats unachievable by mathematics alone. Quantum cryptographic devices typically employ individual photons of light and take advantage of Heisenberg's uncertainty principle, according to which measuring a quantum system in general disturbs it and yields incomplete information about its state before the measurement. Eavesdropping on a quantum communications channel therefore causes an unavoidable disturbance, alerting the legitimate users. Quantum cryptography exploits this effect to allow two parties who have never met and who share no secret information beforehand to communicate in absolute secrecy under the nose of an adversary. Quantum techniques also assist in the achievement of subtler cryptographic goals, important in the post-cold war world, such as enabling two mutually distrustful parties to make joint decisions based on private information, while compromising its confidentiality as little as possible.

CHARLES H. BENNETT, GILLES BRASSARD and ARTUR K. EKERT share a deep interest in the fundamental connections between physics and the theory of computation. Bennett and Brassard have pioneered the field of quantum cryptography. Since 1973 Bennett has been a researcher at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. In 1979 Brassard became a professor of computer science at the Université de Montréal, where he currently holds the E.W.R. Steacie Memorial Fellowship. Ekert, who last year received his doctoral degree from the University of Oxford, is a Junior Research Fellow at Merton College at Oxford. This is Bennett's third article for *Scientific American.*

The art of cryptography began at least 2,500 years ago and has played an important role in history ever since. Perhaps one of the most famous cryptograms, the Zimmermann Note, propelled the U.S. into World War I. When the cryptogram was broken in 1917, Americans learned that Germany had tried to entice Mexico to join its war effort by promising Mexico territories in the U.S.

Around this time Gilbert S. Vernam of American Telephone and Telegraph Company and Major Joseph O. Mauborgne of the U.S. Army Signal Corps developed the first truly unbreakable code called the Vernam cipher [*see box on page 28*]. One distinctive feature of the code is its need for a key that is as long as the message being transmitted and is never reused to send another message. (The Vernam cipher is also known as the one-time pad from the practice of furnishing the key to spies in the form of a tear-off pad, each sheet of which was to be used once and then carefully destroyed.) The discovery of the Vernam cipher did not create much of a stir at the time, probably because the cipher's unbreakability was not definitively proved until later and because its massive key requirements made it impractical for general use.

Because of this limitation, soldiers and diplomats continued to rely on weaker ciphers using shorter keys. Consequently, during World War II, the Allies were able to read most of the secret messages transmitted by the Germans and Japanese. These ciphers, though breakable, were by no means easy to crack. Indeed, the formidable task of breaking increasingly sophisticated ciphers was one of the factors that stimulated the development of electronic computers.

Academic interest in cryptology grew more intense in the mid-1970s, when Whitfield Diffie, Martin E. Hellman and Ralph C. Merkle, then at Stanford University, discovered the principle of public-key cryptography (PKC). Soon afterward, in 1977, Ronald L. Rivest, Adi Shamir and Leonard M. Adleman, then at the Massachusetts Institute of Technology, devised a practical implementation [see "The Mathematics of Public-Key Cryptography," by Martin E. Hellman; SCIENTIFIC AMERICAN, August 1979].

Public-key cryptosystems differ from all previous schemes in that parties wishing to communicate do not need to agree on a secret key beforehand. The idea of PKC is for a user, whom we shall call Alice, to choose randomly a pair of mutually inverse transformations—to be used for encryption and decryption; she then publishes the instructions for performing encryption but not decryption. Another user, Bob, can then use Alice's public-encryption algorithm to prepare a message that only she can decrypt. Similarly, anyone, including Alice, can use Bob's public-encryption algorithm to prepare a message that only he can decrypt. Thus, Alice and Bob can converse secretly even though they share no secret to begin with. Public-key cryptosystems are especially suitable for encrypting electronic mail and commercial transactions, which often occur between parties who, unlike diplomats and spies, have not anticipated their need to communicate secretly.

Offsetting this advantage is the fact that public-key systems have not been

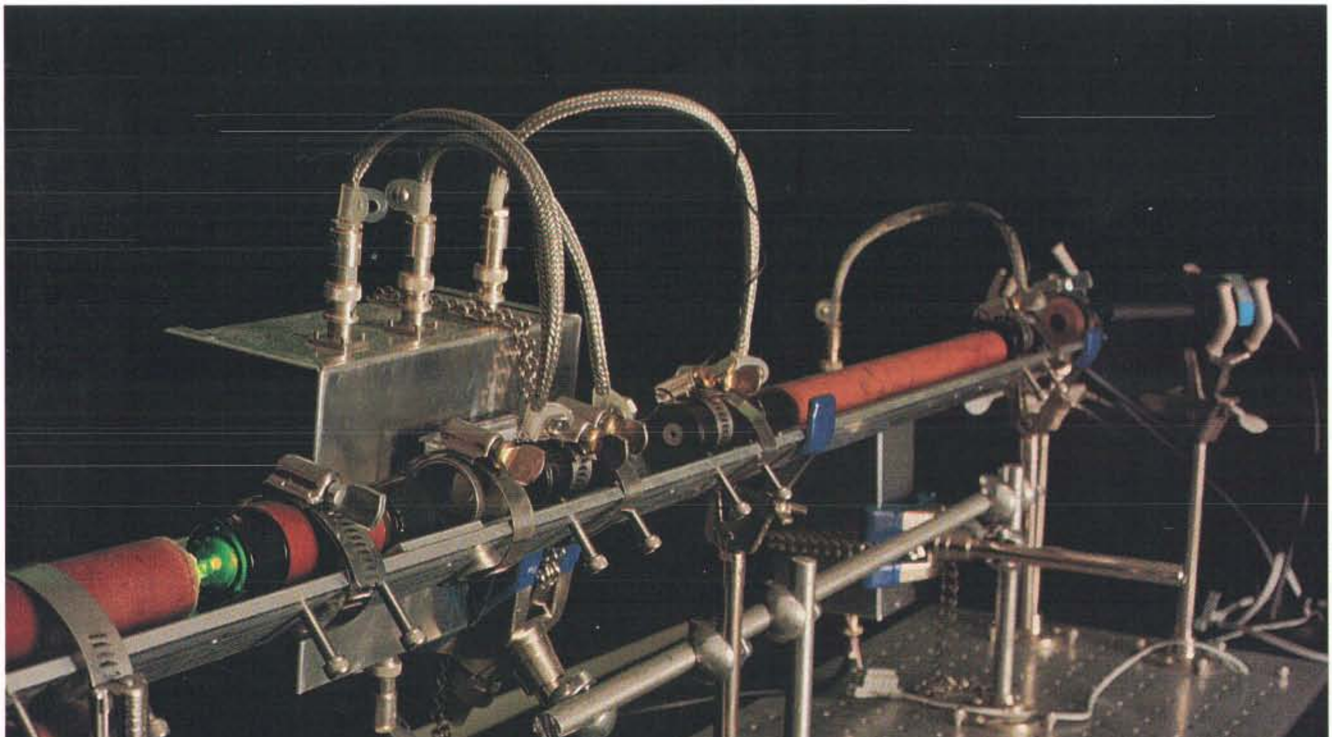
proven to be secure. Indeed, in 1982 Shamir, now at the Weizmann Institute of Science in Israel, cracked one of the early public-key cryptosystems, the knapsack cipher. Poe could be smiling from the grave, knowing there is a clever method of attack, as yet undiscovered, that could break any of these ciphers in a few minutes.

Several years before the discovery of public-key cryptography, another striking development had quietly taken place: the union of cryptography with quantum mechanics. Around 1970 Stephen J. Wiesner, then at Columbia University, wrote a paper entitled "Conjugate Coding," explaining how quantum physics could be used, at least in principle, to accomplish two tasks that were impossible from the perspective of classical physics. One task was a way to produce bank notes that would be physically impossible to counterfeit. The other was a scheme for combining two classical messages into a single quantum transmission from which the receiver could extract either message but not both. Unfortunately, Wiesner's paper was rejected by the journal to which he sent it, and it went unpublished until 1983. Meanwhile, in 1979, two of us (Bennett and Brassard) who knew of Wiesner's ideas began thinking of how to combine them with public-key cryptography. We soon realized that they could be used as a substitute for PKC: two users, who shared no secret initially, could communicate secretly, but now with absolute and provable security, barring violations of accepted physical laws.

Our early quantum cryptographic schemes, developed between 1982 and 1984, were somewhat impractical, but refinements over the next few years culminated in the building of a fully working prototype at the IBM Thomas J. Watson Research Center in 1989. John Smolin, now at the University of California at Los Angeles, helped to build the electronics and optics for the apparatus, and François Bessette and Louis Salvail of the University of Montreal assisted in writing the software. At about the same time, the theoretical ideas of David Deutsch of the University of Oxford led one of us (Ekert) to conceive of a slightly different cryptosystem based on quantum correlations. In early 1991, utilizing ideas conceived by Massimo Palma of the University of Palermo, John Rarity and Paul Tapster of the British Defence Research Agency started experiments implementing Ekert's cryptosystem.

To explain how such systems work, we need to describe in more detail some aspects of the mathematics of classical cryptography, especially the role of the key. In the early days of cryptography the security of a cipher depended on the secrecy of the entire encryption and decryption procedure. Today such procedures are usually known publicly, but the key is kept secret. In such ciphers the key is used to control and customize the encryption and decryption processes in such a way that an adversary who has intercepted the cryptogram and knows the general method of encryption but not the key will not be able to infer anything useful about the original message. Consequently, the cryptogram may be broadcast over a public channel such as a radio or printed in a newspaper. The key, however, must be sent through a very secure private channel, such as a clandestine meeting or a delivery by a trusted courier. Although the distribution of a key over private channels is expensive, it makes possible subsequent secret communication over inexpensive public channels.

Ultimately, the security of a cryptogram depends on the length of the key. In two brilliant papers written in the 1940s, Claude E. Shannon, then at Bell Laboratories, showed that if the key is shorter than the message being encrypted, some information about the message can be inferred from the cryptogram by a sufficiently powerful adversary. This leakage of information occurs regardless of how complicated the en-



QUANTUM DEVICE generates and measures extremely faint flashes of polarized light, providing a secure way to transmit information [*see illustration on pages 32 and 33*]. On average, each flash consists of one tenth of a photon.

The Cipher of Ché Guevara

When in 1967 the Bolivian army captured and executed the revolutionary Ché Guevara, they found on his body a worksheet showing how he prepared a message for transmission to Cuban president Fidel Castro. Guevara used the unbreakable cipher invented by Gilbert Vernam in 1918. The letters of Guevara's message (in Spanish) were first translated into one- and two-digit decimal numbers by a fixed rule, namely:

A 6	E 8	I 39	M 70	Q 71	U 52	Y 1
B 38	F 30	J 31	N 76	R 58	V 50	Z 59
C 32	G 36	K 78	O 9	S 2	W 56	
D 4	H 34	L 72	P 79	T 0	X 54	

By itself this procedure would have provided virtually no protection. The message digits were then strung together in convenient five-digit blocks. They became the top line of each three-line group on the worksheet. The middle line of each group is the key, a sequence of random digits known only to Guevara and Castro.

Next the message and key were added (without carries) to produce a cryptogram, forming the bottom line of each three-line group. Because of the addition of the random key digits, this cryptogram is itself a random decimal sequence, carrying no information about the original message, except to someone who knows the key. The cryptogram was then transmitted to Cuba by an insecure channel such as shortwave radio. At the receiving end, Castro's cipher office would have subtracted the same random key digits, reconstructing the number sequence in the top row, and then would have translated the numbers back into the letters of the message.

Many spies and diplomats have used the Vernam cipher throughout the 20th century. The key, rather than consisting of decimal digits, can be a long random sequence of the binary digits 0 and 1, and the additions and subtractions would be carried out in base 2 by machine, rather than in base 10 by hand. Nevertheless, the key must still be hand-carried from the place where it is generated to the places where it will be used, and it must be assiduously guarded during all phases of delivery and storage to prevent it from falling into the hands of an adversary.

cryption process may be. In contrast, the message can be completely and unconditionally hidden from the eavesdropper by cryptosystems such as the Vernam cipher, in which the key is as long as the message, is purely random and is used only once.

Even the Vernam cipher is only as secure as its key distribution and key storage. Because of the great difficulty of supplying secure new keys for every message, the Vernam cipher is impractical for general commercial uses, although it is routinely employed for diplomatic communications such as those exchanged over the Moscow-Washington hot line. In contrast, the most widely used commercial cipher, the Data Encryption Standard, depends on a 56-bit secret key, which is reused for many encryptions over a certain length of time. This scheme simplifies the problem of secure key distribution and storage, but it does not eliminate it.

A fundamental problem remains. In principle, any classical private channel can be monitored passively, without the sender or receiver knowing that the eavesdropping has taken place. For example, a key carried by a trusted courier might have been read en route by a surreptitious high-resolution x-ray scan or other sophisticated imaging technique without the courier's knowledge. More generally, classical physics—the theory of macroscopic bodies and phenomena such as paper documents, magnetic tapes and radio signals—allows all physical properties of an object to be measured without disturbing those properties. Since all information, including a cryptographic key, is encoded in measurable physical properties of some object or signal, classical theory leaves open the possibility of passive eavesdropping, because it allows the eavesdropper in principle to measure physical properties without disturbing them.

This is not the case in quantum theory, which forms the basis for quantum cryptography. Quantum theory is believed to govern all objects, large and small, but its consequences are most conspicuous in microscopic systems such as individual atoms or subatomic particles. The act of measurement is an integral part of quantum mechanics, not just a passive, external process as in classical physics. So it is possible to design a quantum channel—one that carries signals based on quantum phenomena—in such a way that any effort to monitor the channel necessarily disturbs the signal in some detectable way. The effect arises because in quantum theory, certain pairs of physical prop-

erties are complementary in the sense that measuring one property necessarily disturbs the other. This statement, known as the Heisenberg uncertainty principle, does not refer merely to the limitations of a particular measurement technology: it holds for all possible measurements.

The uncertainty principle can be applied to design a completely secure channel based on the quantum properties of light. The smallest unit, or quantum, of light is the photon, which can be thought of as a tiny, oscillating electric field. The direction of the oscillation is known as the photon's polarization. Ordinary light consists of photons that have many different polarizations. But if the light passes through a polarizing filter, such as those used in sunglasses, only photons having a particular polarization will make it through. Which polarization is transmitted depends on the orientation of the filter. In sunglasses the filters are oriented to transmit vertically polarized light because such light reflects off most horizontal surfaces with less glare. But if the glasses are turned 90 degrees, so that one lens is directly above the other, they will transmit horizontally polarized light, augmenting the glare instead of diminishing it.

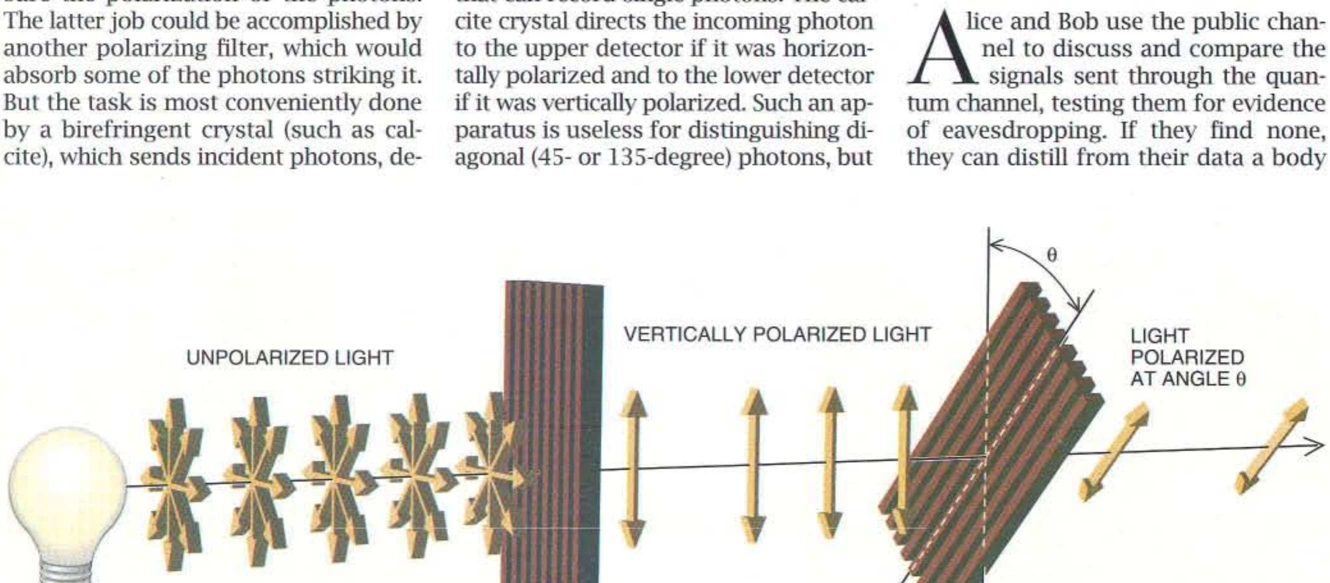
To construct a quantum channel, one needs a polarizing filter or other means for the sender to prepare photons of selected polarizations and a way for the receiver to measure the polarization of the photons. The latter job could be accomplished by another polarizing filter, which would absorb some of the photons striking it. But the task is most conveniently done by a birefringent crystal (such as calcite), which sends incident photons, depending on their polarization, into one of two paths without absorbing any [*see illustration on next page*].

A photon encountering a calcite crystal behaves in one of two ways depending on its polarization in relation to the crystal. The photon may pass straight through the crystal and emerge polarized perpendicular to the crystal's optic axis, or it may be shifted and emerge polarized along that axis. If the photon entering the crystal is already polarized in one of these two directions, it will undergo no change of polarization but will be deterministically routed into the straight or shifted path, respectively. If a photon polarized at some intermediate direction enters the crystal, however, it will have some probability of going into each beam and will be repolarized according to which beam it goes into, forgetting its original polarization. The most random behavior occurs when the photon is polarized halfway between these two directions, that is, at 45 or 135 degrees. Such photons are equally likely to go into either beam, revealing nothing about their original polarization and losing all memory of it.

Suppose Bob is told in advance that a given photon is polarized in one of the two "rectilinear" directions, vertical (90 degrees) or horizontal (0 degrees) without being informed of the specific polarization. Then he can reliably tell which direction by sending the photon into an apparatus consisting of a vertically oriented calcite crystal and two detectors, such as photomultiplier tubes, that can record single photons. The calcite crystal directs the incoming photon to the upper detector if it was horizontally polarized and to the lower detector if it was vertically polarized. Such an apparatus is useless for distinguishing diagonal (45- or 135-degree) photons, but these can be reliably distinguished by a similar apparatus that has been rotated 45 degrees from the original orientation. The rotated apparatus, in turn, is useless for distinguishing vertical from horizontal photons. According to the uncertainty principle, these limitations apply not just to the particular measuring apparatus described here but to any measuring device whatsoever. Rectilinear and diagonal polarizations are complementary properties in the sense that measuring either property necessarily randomizes the other.

We can now describe the simple scheme for quantum key distribution that two of us (Bennett and Brassard) proposed in 1984 and that we dubbed "BB84." The purpose of the scheme is for Alice and Bob to exchange a secret random key that they can subsequently use, as in the Vernam cipher, to send meaningful secret messages when the need arises. Like other quantum key distribution schemes, BB84 uses a quantum channel, through which Alice and Bob send polarized photons, in conjunction with a classical public channel, through which they send ordinary messages. An eavesdropper, whom we shall call Eve, is free to try to measure the photons in the quantum channel, but, as noted above, she cannot in general do this without disturbing them. Furthermore, she learns the entire contents of messages sent through the public channel but assume for the moment that she could not disturb or alter these messages even if she wanted to.

Alice and Bob use the public channel to discuss and compare the signals sent through the quantum channel, testing them for evidence of eavesdropping. If they find none, they can distill from their data a body



UNPOLARIZED LIGHT enters a filter, which absorbs some of the light and polarizes the remainder in the vertical direction. A second filter tilted at some angle absorbs some of the polarized light and transmits the rest, giving it a new polarization.

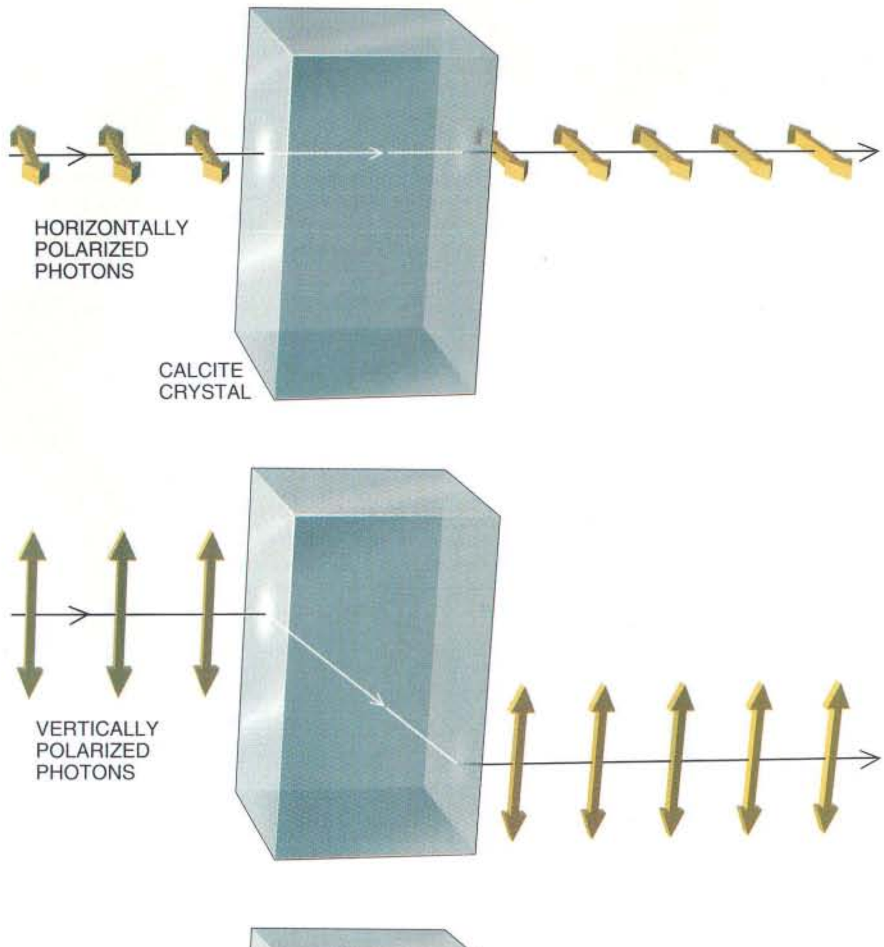
of information that is certifiably shared, random and secret, regardless of Eve's technical sophistication and the computing power at her disposal [*see box on opposite page*]. The scheme works as follows:

First, Alice generates and sends Bob a sequence of photons whose polarizations she has chosen at random to be either 0, 45, 90 or 135 degrees. Bob receives the photons and, for each photon, decides randomly whether to measure its rectilinear or diagonal polarization.

Next Bob announces publicly, for each photon, which type of measurement he has made (rectilinear or diagonal) but not the measurement result (for example, 0, 45, 90 or 135 degrees). Alice tells him publicly, for each photon, whether he has made the right kind of measurement. Alice and Bob then discard all cases in which Bob has made the wrong measurement or in which his detectors have failed to register a photon at all (existing detectors are not 100 percent efficient). If no one has eavesdropped on the quantum channel, the remaining polarizations should be shared secret information between Alice and Bob.

Alice and Bob next test for eavesdropping, for example, by publicly comparing and discarding a randomly selected subset of their polarization data. If the comparison test shows evidence of eavesdropping, Alice and Bob discard all their data and start over with a fresh batch of photons. Otherwise they adopt the remaining polarizations, which have never been publicly mentioned, as shared secret bits, interpreting horizontal or 45-degree photons as binary 0's and vertical or 135-degree photons as binary 1's.

Because of the uncertainty principle, Eve cannot measure both rectilinear and diagonal polarizations of the same photon. If, for a particular photon, she makes the wrong measurement, then, even if she resends Bob a photon consistent with the result of her measurement, she will have irretrievably randomized the polarization originally sent by Alice. The net effect is to cause errors in one quarter of the bits in Bob's data that have been subjected to eavesdropping.



CALCITE CRYSTAL can be used to distinguish between horizontally and vertically polarized photons. Horizontally polarized photons pass straight through, whereas vertically polarized photons are deflected. When diagonally polarized photons enter the crystal, they are repolarized at random in either the vertical or horizontal direction and are shifted accordingly.

Directly comparing selected bits to test for errors, as described above, is not very efficient: too many bits must be sacrificed in order to build reasonable confidence that Alice's and Bob's data are identical, especially if the eavesdropping has been rather infrequent, resulting in only a few errors. A much better idea is for Alice and Bob to compare the "parity"—evenness or oddness—of a publicly agreed-on random subset containing about half the bits in their data. For instance, Alice could tell Bob, "I looked at the 1st, 3rd, 4th, 9th,... 996th and 999th of my 1,000 bits of data, and they include an even number of 1's." Bob would then count the number of 1's that he has in those same locations. If he finds an odd number of 1's, he can immediately conclude that his data are different from Alice's. It can be shown that if Alice's and Bob's data differ, comparing the parity of a random subset will detect that fact with a probability of $^1/_2$, regardless of the number and location of errors. It suffices to repeat the test 20 times, with 20 different random subsets, to reduce the chance of an undetected error to less than one in a million.

The BB84 scheme was modified to produce a working quantum cryptography apparatus at IBM [*see illustration on pages 32 and 33*]. The modifications were necessary to deal with practical problems such as noise in the detec-

tors and the fact that the prototype uses dim flashes of light instead of single photons.

The quantum channel, with Alice's sending apparatus at one end and Bob's receiving apparatus at the other, is housed in a light-tight box. During operation, the system is controlled by a personal computer, which contains separate software representations of Alice, Bob and, optionally, Eve.

The leftmost part of Alice's sending apparatus consists of a green light-emitting diode, a lens, a pinhole and filters that provide a collimated beam of horizontally polarized light. Next, electro-optic devices known as Pockels cells are used to change the original horizontal polarization to any of four standard polarization states under Alice's control. This has the same effect as mechanically rotating her polarizing filter into four different positions, but the process can be done much faster.

At the other end of the quantum channel, Bob's receiving apparatus contains a similar Pockels cell, which allows him to choose the type of polarization he will measure, again without actually rotating his detecting device. After the beam passes through Bob's Pockels cell, it is split by a calcite prism into two perpendicularly polarized beams, which are directed into two photomultiplier tubes for the purpose of detecting individual photons.

The sending and receiving apparatuses in the prototype are only about 30 centimeters apart—chiefly because of the desire to keep the device of reasonable dimensions for a desktop—but nothing in principle prevents the technique from being used over much greater distances. For example, quantum transmissions could be sent through several kilometers of optical fiber. If cost and inconvenience were no concern, quantum transmissions could be sent over arbitrarily great distances with negligible losses through an evacuated straight pipe. But quantum key distribution has to compete with classical techniques, which are much cheaper over long distances and may be sufficiently secure.

Recall that the BB84 scheme encodes each bit in a single polarized photon. In contrast, the prototype encodes each bit in a dim flash of light. This introduces a new eavesdropping threat to the system: if Eve taps into the beam by a device such as a half-silvered mirror, she will be able to split each flash into two flashes of lesser intensity, reading one herself while letting the other pass to Bob, its polarization undisturbed. If Eve diverts only a modest fraction of the beam, Bob may not notice the weakening of his signal, or he may attribute it to natural losses in the channel. This attack can be effectively thwarted, at the cost of reducing the rate of data transmission through the quantum channel, by having Alice send very dim flashes—that is, an intensity of less than one photon per flash on average. Such extremely dim flashes can easily be made by filtering out almost all the intensity of a bright flash.

When low-intensity flashes are used, Bob's chance of detecting a photon in any given flash is of course proportionately reduced, but the chance of Bob and Eve both detecting photons in the same flash is reduced even more, as the square of the intensity. The actual apparatus generates intensities of about one tenth of a photon per flash. On the other hand, if Alice's flashes were much brighter (say, thousands of photons per flash), they would be easy prey for the beam-splitting attack. By splitting off only a small fraction of the intensity, Eve could still obtain enough photons from each flash to perform both rectilinear and diagonal measurements and so determine the polarization unambiguously. In other words, the brighter Alice's flashes, the more they behave like classical signals, about which an eavesdropper can gain complete information while introducing negligible disturbance.

Another practical problem arises from the fact that available detectors sometimes produce a response even when no photon has arrived. Such "dark counts" and other imperfections in the

Quantum Key Distribution

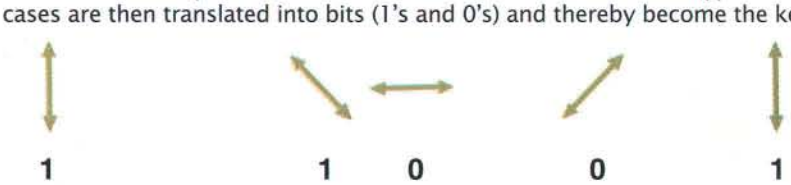
A quantum cryptographic system will allow two people, say, Alice and Bob, to exchange a secret key. The system includes a transmitter and a receiver. Alice uses the transmitter to send photons in one of four polarizations: 0, 45, 90 or 135 degrees. Bobs uses the receiver to measure the polarization. According to the laws of quantum mechanics, the receiver can distinguish between rectilinear polarizations (0 and 90), or it can quickly be reconfigured to discriminate between diagonal polarizations (45 and 135); it can never, however, distinguish both types. The key distribution requires several steps. Alice sends photons with one of four polarizations, which she has chosen at random.

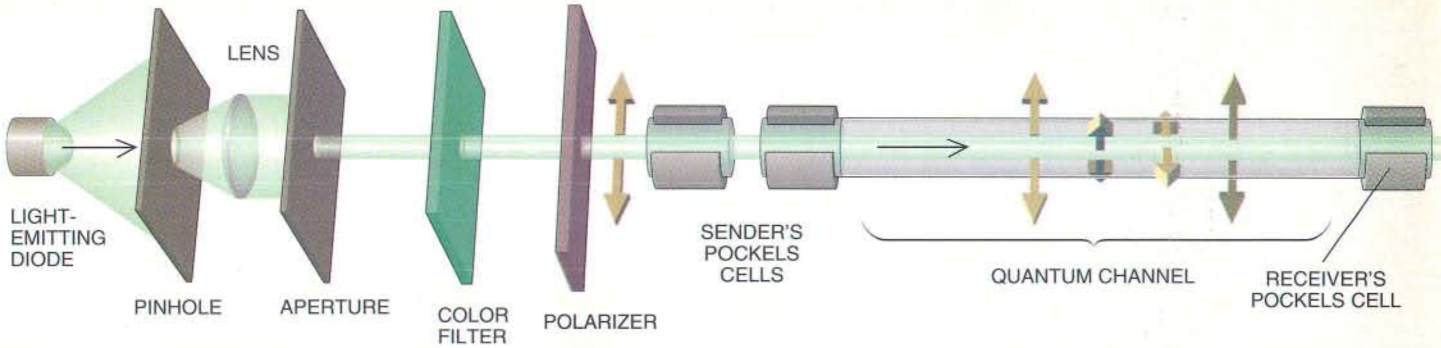
For each photon, Bob chooses at random the type of measurement: either the rectilinear type (+) or the diagonal type (×).

Bob records the result of his measurement but keeps it a secret.

Bob publicly announces the type of measurements he made, and Alice tells him which measurements were of the correct type.

Alice and Bob keep all cases in which Bob measured the correct type. These cases are then translated into bits (1's and 0's) and thereby become the key.





QUANTUM SYSTEM can distribute information in perfect secrecy. The transmitter produces faint flashes of green light from a light-emitting diode. The pinhole, lens and filter create a collimated beam of dim flashes. The light is then polarized horizontally. Two Pockels cells change the polarization to 0, 45, 90 or 135 degrees. The polarized light flashes are released from the transmitter and eventually reach the receiver. There another Pockels cell shifts the polarization by either 45 degrees or not at all. The action of this Pockels cell allows the receiver to choose between measuring rectilinear or diagonal

apparatus lead to errors even when there has been no eavesdropping and make it impractical for Alice and Bob simply to reject their data whenever they find an error in them, as is done in the ideal BB84 protocol. If Bob and Alice find a small number of errors, they must devise a way to correct them and proceed. On the other hand, if they find a large number, indicating significant eavesdropping, they must reject their data and start over.

A variety of techniques are available for Alice and Bob to correct a small number of errors through public discussion, such as the use of error-correcting codes. But these techniques potentially leak information to Eve, who may be listening to the public discussion. Therefore, after the quantum transmission and the error-correcting discussion, Alice and Bob find themselves with what might be thought of as an impure key, a shared body of data that is only partly secret. Information on that key may have leaked to Eve at several stages. She may have gained information by splitting some flashes, by directly measuring others (she cannot do this too often, as it causes errors in Bob's data) and by listening to the public discussion between Alice and Bob. Fortunately, Alice and Bob, because they know the intensity of the light flashes and the number of errors found and corrected, can estimate how much information might have leaked to Eve through all these routes.

In itself, such an impure key is almost worthless. If it were used as a key for the Vernam cipher, for example, it might prove very insecure if the most important part of the message happened to coincide with a part of the key the eavesdropper knew. Fortunately, two of us (Bennett and Brassard), in collaboration with Jean-Marc Robert (then a student of Brassard), developed a mathematical technique known as privacy amplification. Using this technique, Alice and Bob, through public discussion, can take such a partly secret key and distill from it a smaller amount of highly secret key, of which the eavesdropper is very unlikely to know even one bit. The essential idea of privacy amplification is for Alice and Bob, after the eavesdropping has taken place, to choose publicly a length-reducing transformation to apply to their impure key so that partial information about the input conveys almost no knowledge of the output.

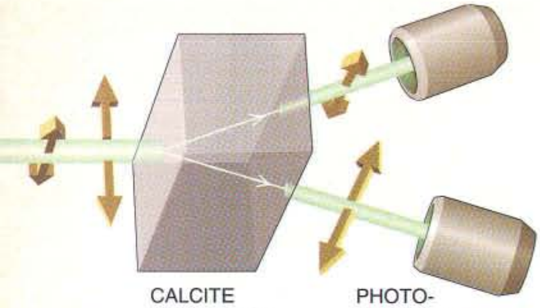
For example, if the input consists of 1,000 bits about which Eve knows at most 200 bits, Alice and Bob can distill nearly 800 highly secret bits as output. Fairly simple techniques can be shown to suffice, and Alice and Bob do not even need to know which partial information the eavesdropper might have about the input in order to choose a function about whose output Eve has almost no information. In particular, it suffices for Alice and Bob to define each bit of the output as the parity of an independent, publicly agreed-on random subset of the input bits, very much as they had done to gain high confidence that their raw quantum data were identical (except that now they keep the parity secret instead of publicly comparing it).

The problem of key security is not entirely solved by secure key distribution. Another weak point is key storage. Once Alice and Bob have established the key, they must store it until it is needed. But the longer they keep the key in, say, their secret safe, the more vulnerable it is to unauthorized inspection. Although principles of engineering may make the safe difficult to crack, the laws of physics always provide the possibility of a security breach. Surprisingly, a cryptosystem can be designed that can guarantee the security of both key distribution and storage by employing quantum correlations. The cryptosystem is based on David Bohm's version of the famous Einstein-Podolsky-Rosen (EPR) effect [see "The Quantum Theory and Reality," by Bernard d'Espagnat; SCIENTIFIC AMERICAN, November 1979].

The EPR effect occurs when a spherically symmetric atom emits two photons in opposite directions toward two observers, Alice and Bob. The two photons are produced in an initial state of undefined polarization. But because of the symmetry of the initial state, the polarizations of the photons, when measured, must have opposite values, provided that the measurements are of the same type. For example, if Alice and Bob both measure rectilinear polarizations, they are each equally likely to record either a 0 (horizontal polarization) or a 1 (vertical), but if Alice obtains a 0, Bob will certainly obtain a 1, and vice versa.

The unusual and important aspect of the EPR effect is that the polarization of both photons is determined as soon as, but not before, one of the photons is measured. This happens no matter how far apart the photons may be at the time. This "classical" explanation of the EPR effect is somewhat counterintuitive, and indeed all classical explanations of the EPR effect involve some implausible element, such as instantaneous action at a distance. Yet the mathematical formalism of quantum mechanics accounts for the EPR effect in a straightforward manner, and experiments have amply confirmed the existence of the phenomenon.

Employing the EPR effect, one of us (Ekert) recently devised a cryptosystem that guarantees the security of both



polarization. In the rectilinear case, a horizontally polarized photon will be directed toward the right photomultiplier; a vertically polarized photon will be directed toward the left photomultiplier.

key distribution and key storage. In a simplified version of this system, as described by N. David Mermin of Cornell University, Alice generates a number of EPR photon pairs, keeping one member of each pair for herself while sending the other to Bob. Alice and Bob measure some of their photons immediately to test for eavesdropping but store the remainder without measuring them. Then, just before the key is needed, they measure and compare some of the stored photons. If no one has tampered with the photons in storage, Bob will always obtain 1 when Alice obtains 0, and vice versa. If no discrepancies are found, Alice and Bob measure the remaining stored photons to obtain the desired key.

Although this procedure works in theory, it cannot be used in practice because of the technical infeasibility of storing photons for more than a small fraction of a second. Thus, at present, the EPR effect is not a practical means for certifying the security of key storage.

Although the best-known application of cryptography is secret communication, two other applications are probably more important in peacetime. The first is authentication: certifying that a message is from whom it says it is from and has not been altered in transit. The second is the task of maintaining the confidentiality of private information while it is being used to reach public decisions.

For most of recorded history, authentication has depended on physical objects that are hard to copy, such as seals and signatures. Such devices provide limited security, and they cannot be used at all for digital electronic documents, such as bank transactions, which are often transmitted over insecure telecommunications lines.

Fortunately, several mathematical techniques are available for authenticating digital messages. In 1979 Mark N. Wegman and J. Lawrence Carter of IBM discovered a digital authentication scheme that does provide provable security. Like the Vernam cipher, it requires that the sender and receiver possess beforehand a shared secret key, part of which is used up each time a message is authenticated.

Wegman-Carter authentication and quantum key distribution can each benefit the other. On the one hand, the quantum technique supplies the secret key bits consumed by the authentication method. On the other, Wegman-Carter authentication can be used to conduct quantum key distribution successfully even in the presence of a more powerful adversary, that is, someone who could alter the public channel messages as well as listening to them.

Quantum cryptography may also prove useful for protecting private information while it is being used to make public decisions. The classic example of such discreet decision making is the "dating problem," in which two singles seek a way of making a date if and only if each likes the other, without disclosing any further information. For example, if Alice likes Bob, but Bob does not like Alice, the date should be called off without Bob's finding out that Alice likes him (on the other hand, it is logically unavoidable for Alice to learn that Bob does not like her, because if he did the date would be on).

There are many other situations in which joint decision making between corporate or governmental organizations or between an individual and an organization depends on confidential information that the negotiating parties would rather not disclose in full [see "Achieving Electronic Privacy," by David Chaum; SCIENTIFIC AMERICAN, August]. A discreet solution to the dating problem, or any other joint decision depending on private data, can be obtained by having Alice and Bob confide their private data to a trusted intermediary (Eve, for example) and letting her make the decision. The hazards of this approach are obvious: Alice and Bob must trust Eve both to make the decision correctly in the first place and never to reveal the private data.

Various other techniques allow public decisions based on private inputs to be made without a single trusted intermediary. For example, a successful protocol can be set up among many participants that will fail only if a majority of the participants conspire to spoil the output or reveal the inputs. On the other hand, if two parties believe in the security of public-key cryptosystems, they can make decisions discreetly without any intermediary. In 1982 Andrew C.-C. Yao, then at Stanford University, was one of the first to investigate this problem.

Recently Claude Crépeau of the École Normale Supérieure and CNRS and his student Marie-Hélène Skubiszewska, in collaboration with two of us (Bennett and Brassard), have shown that a quantum apparatus similar to the one already built for key distribution can be used to make joint decisions discreetly without intermediaries and without unproved mathematical assumptions. Discreet decision making can be implemented by repeated application of a curious information-processing procedure called oblivious transfer. The procedure is a version of Wiesner's feat of sending two messages in such a way that the receiver can read either one but not both. In 1981 Michael O. Rabin of Harvard University formalized the notion of oblivious transfer, unaware of Wiesner's pioneering but unpublished work a decade earlier. Later, Crépeau, Joe Kilian, then at the Massachusetts Institute of Technology, and others demonstrated that oblivious transfer could be applied to discreet decision making.

One particularly attractive feature of discreet quantum decision making is that, unlike key distribution, it is worthwhile even over short distances. Unfortunately, known implementations are mathematically rather inefficient, requiring many thousands of photons to be sent and received to reach even simple decisions. If its mathematical efficiency can be sufficiently improved, discreet decision making may become the most important application of quantum cryptography.

FURTHER READING

THE CODEBREAKERS: THE STORY OF SECRET WRITING. David Kahn. Macmillan, 1967.

MODERN CRYPTOLOGY: A TUTORIAL. Gilles Brassard in *Lecture Notes in Computer Science*, Vol. 325. Springer-Verlag, 1988.

QUANTUM CRYPTOGRAPHY BASED ON BELL'S THEOREM. Artur K. Ekert in *Physical Review Letters*, Vol. 67, No. 6, pages 661–663; August 5, 1991.

EXPERIMENTAL QUANTUM CRYPTOGRAPHY. Charles H. Bennett, François Bessette, Gilles Brassard, Louis Salvail and John Smolin in *Journal of Cryptology*, Vol. 5, No. 1, pages 3–28; 1992.

QUANTUM CRYPTOGRAPHY WITHOUT BELL'S THEOREM. Charles H. Bennett, Gilles Brassard and N. David Mermin in *Physical Review Letters*, Vol. 68, No. 5, pages 557–559; February 3, 1992.

Mountain Sickness

The varied and subtle symptoms of this potentially lethal disorder humble many who scale the summits. But the problem is often preventable

by Charles S. Houston

On New Year's Eve in 1960, I received a call from a young man who had left his sick companion high in the mountains near Aspen, Colo., to ski out for help. A rescue party was assembled, and the following evening the patient was brought to a hospital for treatment of what I anticipated from his friend's description of the problem would be pneumonia. Instead the young man had an accumulation of fluid in both lungs that was unlike the inflammation seen in pneumonia.

Since this finding was unusual, a distinguished cardiologist urged me to publish my observations. The resulting short article in the *New England Journal of Medicine* brought hundreds of letters describing similar altitude-related illnesses that had been characterized as pneumonia. The discovery of what seemed to be a new form of mountain sickness—subsequently named high-altitude pulmonary edema—quickened my long-standing interest in mountain climbing and the medical problems associated with high altitudes. In 1936, as a medical resident, I had accompanied a team of climbers to the Himalayas where I had seen, but failed to recognize, my first case of altitude sickness.

High-altitude pulmonary edema and other symptoms of illness produced by high altitudes cause significant medical problems. Today record numbers of people visit mountains to climb, ski, walk or simply vacation. In recent years rapid transportation and better equipment have made mountain climbing more popular. Between 1903 and 1912, for instance, only 42 climbers attempted to reach the summit of 20,400-foot Mount McKinley in Alaska. None was successful. Between 1988 and 1990, 2,923 attempted the climb; 1,659 made it. Accordingly, the number of victims of high-altitude pulmonary edema and associated illnesses has increased. Fourteen percent of the fatalities on Mount McKinley have been attributed to altitude sickness.

Mountain sickness is caused primarily by a lack of oxygen, or hypoxia. Atmospheric pressure decreases as one moves away from sea level, and because the percentage of oxygen in air remains constant, the concentration of oxygen is decreased. Lower levels of oxygen initiate a series of potentially fatal physiological changes.

Yet prolonged exposure to altitude or, conversely, to sea-level conditions that cause a lack of oxygen may produce adjustments called acclimatization. This process enables people to survive with levels of oxygen that would otherwise cause serious problems. Understanding the many dangers of hypoxia, the different forms of altitude illness and the process of acclimatization is crucial to safety on the stunning mountains that crown the earth.

Although only recently defined as a form of altitude sickness, pulmonary edema at high altitudes is far from new. In A.D. 403 a Chinese archivist, Hui Jiao, traveled the Silk Route and recorded his traveling companion's illness: "The wind was chilling to the bones on the shady north side of the Lesser Snowy Mountains. Hui Jing was in a serious condition, frothing at the mouth, losing his strength rapidly and fainting away now and then. Finally he dropped dead on the snowy ground."

One hundred and fifty years later Mogul chieftain Mirza Muhammad Haidar described the often devastating condition that affected his troops campaigning on the Tibetan Plateau (at 13,000 to 15,000 feet). The symptoms ranged from weakness and shortness of breath to hallucinations, culminating in coma and often death. At the end of the 16th century, Jesuit priest José Acosta wrote a similar description while he was crossing a high pass in the Andes.

Hui Jiao, Haidar and Acosta could not know what caused this illness, because the nature of the earth's atmosphere was not understood until the mid-17th century. At that time, a series of experiments conducted by Gaspar Berti, Evangelista Torricelli and Florin Périer showed that the atmosphere did have weight and that its pressure decreases as the altitude increases.

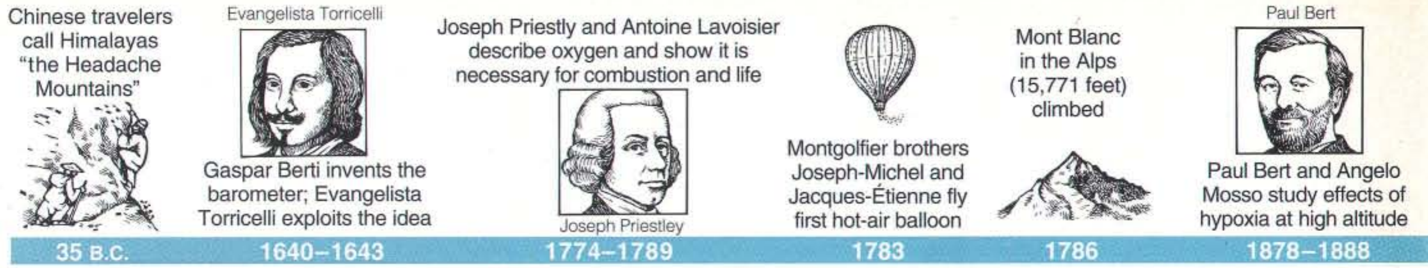
In the 18th century people began climbing mountains for pleasure, and physicians started to define clinically the physiological effects of lowered atmospheric pressure. In 1786 climbers ascended Mont Blanc, the highest peak in Europe at 15,771 feet. During the next decade, Horace-Bénédict de Saussure described how his heart and respiration were affected by altitude. Soon oxygen was identified, and its importance for supporting life as well as combustion was demonstrated. But the connection between hypoxia and mountain sickness remained unrecognized.

The golden age of alpine climbing began in the 1850s, when British doctor Albert Smith undertook a career lecturing about his ascent of Mont Blanc. Thousands of people went into the mountains, and in the next three decades all the alpine summits were reached. A few explorers turned to the Himalayas and the Andes. Balloons, first flown a century before, also became fashionable, and the daring went high enough to suffer from altitude.

CHARLES S. HOUSTON is a physician who has loved mountains ever since his first alpine ascent in 1925. In 1936 he climbed in the Himalayas, and in 1938 he led the first American expedition to K2 (at 28,251 feet). After receiving his M.D. from Columbia University, Houston entered the U.S. Navy, where he served as a flight surgeon, specializing in high altitude. Houston, who taught internal medicine at the University of Vermont, continues his altitude-related research.

MOUNTAINEER tackles Mount Fitzroy (at 11,070 feet) in Patagonia. By ascending slowly and allowing for acclimatization, experienced climbers can avoid some of the perils of mountain sickness.

Significant Events in the History of Mountain Climbing



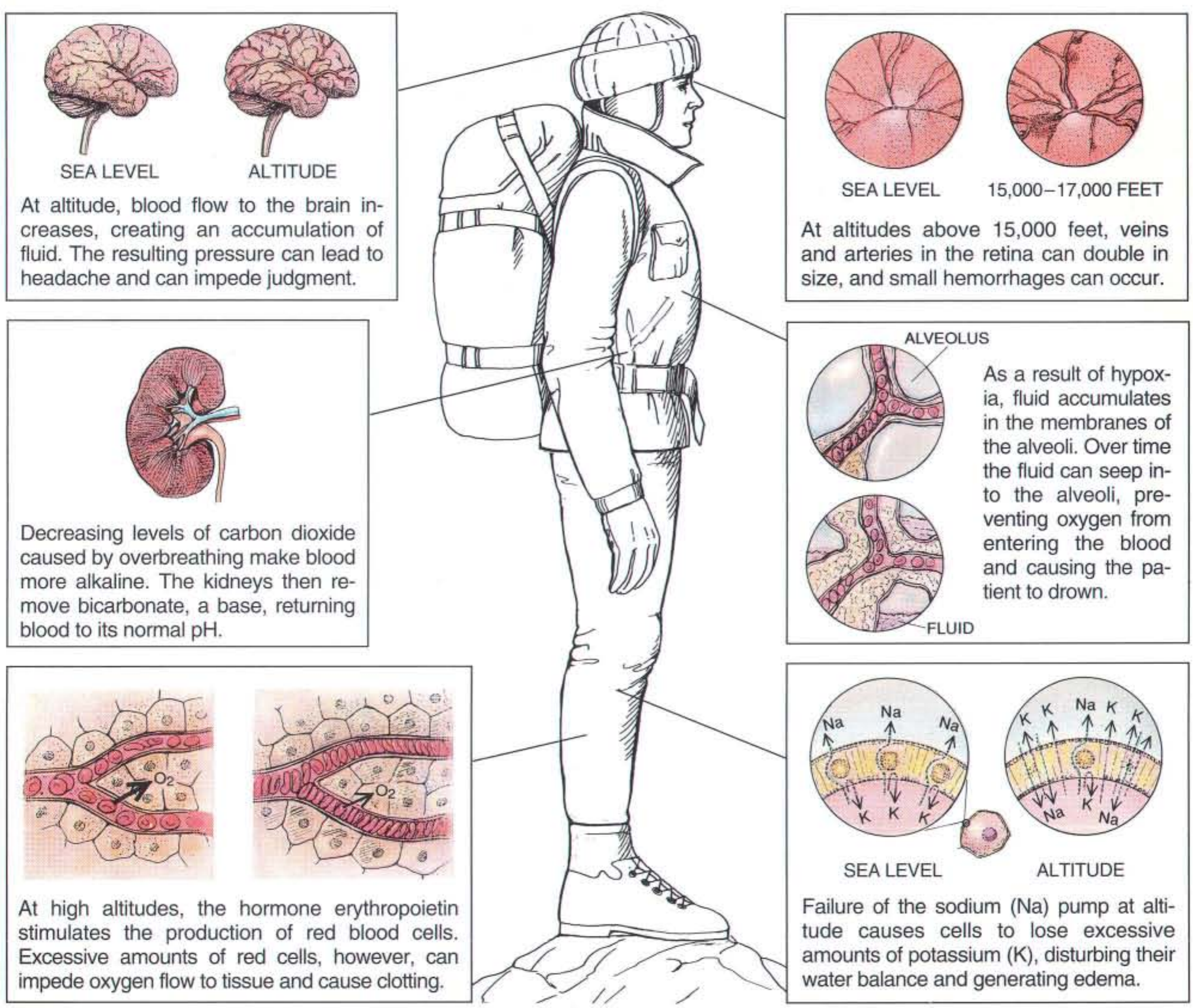
The words "mountain sickness" began to appear in the popular and medical literature. Stories, often exaggerated, described the symptoms that plagued climbers and some animals, but there was little agreement about their cause. Failure to mention the miserable effects often raised doubts that a summit had been reached.

In the last decades of the 19th century, the work of two physicians clarified the relation between thin air and mountain sickness. Paul Bert conducted studies at simulated altitudes in an iron decompression chamber and measured the carriage of oxygen by hemoglobin. After showing that blood contained less oxygen at altitude than it did at sea level, he took himself to a simulated 21,000 feet—while breathing oxygen from a leather bag. Because no symptoms appeared, he concluded that lack of oxygen, rather than lack of pressure, caused mountain sickness.

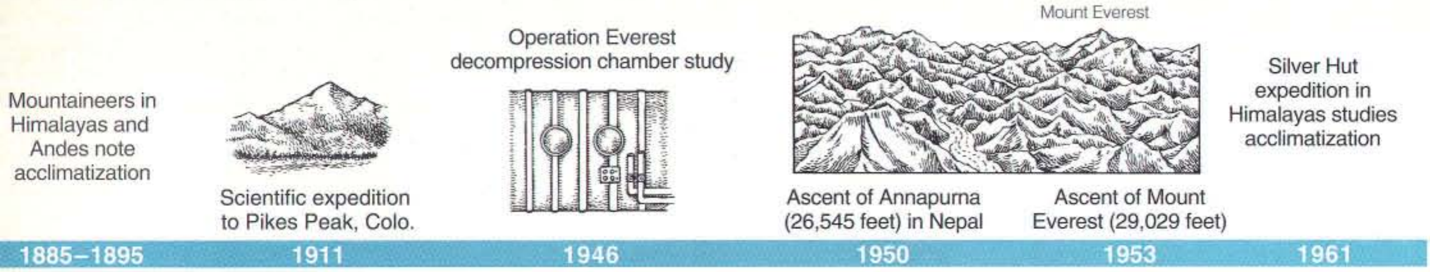
Bert's contemporary, Angelo Mosso, studied men on top of 15,203-foot Monte Rosa in Italy as well as in a decompression chamber. He too decided that hypoxia caused mountain sickness but argued that the lack of carbon dioxide resulting from overbreathing was more important.

The imperatives of aviation during

Some of the Physiological Effects of Hypoxia



and the Understanding of Altitude Sickness



the two world wars and a growing interest in mountain sports increased people's understanding of hypoxia. Studies showed that there are several different forms of mountain sickness—one an unpleasant, but minor problem, others more serious and potentially fatal. Symptoms depend on which parts of the body are most susceptible.

The brain is extremely vulnerable. On average, the human brain receives 10 to 15 percent of the heart's output and uses 15 to 20 percent of the oxygen consumed by the body. The cerebral cortex, where the most complex mental activity takes place, is the most demanding region. It is not surprising, then, that hypoxia affects the higher centers of the brain first, including judgment. Indeed, the effects of hypoxia resemble those of alcohol.

Headache is the most prominent symptom of mountain sickness, but its cause remains unclear. One explanation may lie in the response of a sensitive membrane called pia mater, which covers brain tissue and blood vessels. As arterial oxygen levels fall, blood flow to the brain increases. Consequently, distended vessels or swollen brain tissue may press on the surrounding membranes, causing headache. At the same time, the lack of oxygen stimulates increased breathing, which pumps carbon dioxide out of the lungs and blood. Lowered carbon dioxide, in turn, causes a decrease in blood flow to the brain.

Whether blood flow to the brain ultimately increases or decreases depends on the balance between hypoxia and hypocapnia (decreased carbon dioxide) as well as on the sensitivity of the receptors that govern blood flow. This sensitivity varies from individual to individual, which may explain why reports of the severity of headache differ so greatly and why studies of cerebral blood flow at altitude often produce contradictory results. The nausea, vomiting and disturbed sleep so typical of mountain sickness may be the result of altered blood flow to the midbrain, where these functions are controlled.

Hearing, smell and taste are not affected by altitude, but appetite usually decreases, in time leading to weight loss. It is unclear whether such loss results from malabsorption or simply from a smaller caloric intake.

Clearly, the effects of hypoxia are quite varied. The signs and symptoms differ depending on the altitude reached, the speed of ascent and other influences. What may seem a rather unpleasant but mild illness can change rapidly to one that is life-threatening. Not uncommonly, for example, a visitor to a mountain resort will feel a headache and general malaise and will soon develop shortness of breath and a cough. These symptoms can lead to coma or hallucinations and, if not cared for properly, death.

Although these symptoms form a continuum, for convenience I will describe them as separate entities. The most common form of illness is called acute mountain sickness, and it affects a quarter of the visitors to mountain resorts. The signs are usually those described above: headache, nausea, vomiting, anorexia and lassitude. Seldom fatal, acute mountain sickness can become serious.

The affliction of my patient in Aspen, high-altitude pulmonary edema, is less frequent but more dangerous than acute mountain sickness. Studies have shown that most people going to even moderate altitudes, such as 8,000 feet, develop some fluid in the tissue that separates the alveoli, or air sacs, from the capillary blood vessels. This fluid is usually promptly reabsorbed.

If the fluid accumulates within the alveoli, however, passage of oxygen from lung to blood is impaired. At this point, hypoxia worsens, and more fluid leaks into the air sacs. The victim can literally drown in his or her own juices. The symptoms are increasing shortness of breath and an irritating cough that produces frothy, blood-tinged sputum.

More serious than high-altitude pulmonary edema is high-altitude cerebral edema. In this form of sickness, which can occur at heights as low as 9,000 feet, the central nervous system is affected: parts of the brain become waterlogged. Early warning signs include ataxia, manifested by a staggering drunken walk or by difficulty in performing fine-motor skills. These changes are attributed to swelling of the cerebellum, the area of the brain that controls balance. Mental confusion and hallucinations are also common. If untreated, cerebral edema can cause death.

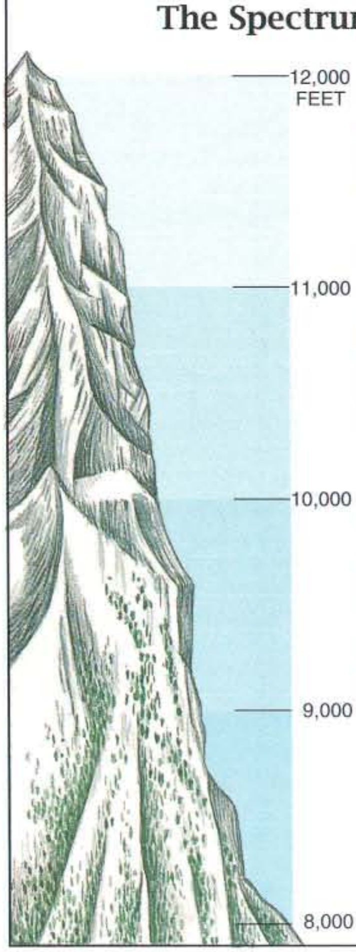
At the end of the physiological continuum lies chronic mountain sickness, an uncommon problem affecting a few people who live permanently above 12,000 feet. Victims experience fatigue, palpitations, chest pain and swelling of the ankles, and they develop an overabundance of red cells and blood clots in their veins and lungs. Going to low altitude corrects the often fatal affliction.

Above heights of 10,000 feet, changes can also occur in the eyes. Because the large supply of oxygen that the rods require is no longer readily available, vision in dim light is decreased by 50 percent. In addition, above 14,000 feet people may develop tiny bleeding spots in back of their eyes. These retinal hemorrhages are usually unnoticed. Researchers have been unable to determine whether the hemorrhages have prognostic or diagnostic importance. Some believe they reflect bleeding elsewhere in the body, a rather unattractive thought for climbers. Those who contend that repeated or prolonged exposure to severe hypoxia may leave permanent brain damage blame it on similar bleeding in the brain.

Intricate, seemingly contradictory changes take place as the body responds to hypoxia. The depth and rate of breathing increase, bringing more air deep into the lungs, raising alveolar oxygen pressure and reducing blood levels of carbon dioxide. This response, however, creates a crisis. Hypoxia can and must be relieved by hyperventilation. Yet the pH of the body must also be maintained. Because this balance is brought about by preserving a certain concentration of carbon dioxide, hyperventilation threatens homeostasis.

Faced with this dilemma, the body compromises by increasing respiration enough to elevate alveolar oxygen and by excreting bicarbonate to offset the alkalosis caused by dwindling blood levels of carbon dioxide. Success in balancing these two conflicting demands determines the effectiveness of acclimatization as well as whether or not an in-

The Spectrum of Altitude Sickness



Chronic mountain sickness affects people who lose their tolerance to high altitude or who fail to acclimatize. It is characterized by fatigue and chest pain as well as by an increase in red blood cell count and, sometimes, heart failure. Chronic mountain sickness can be alleviated by descent to sea level.

High-altitude cerebral edema can occur at 9,000 feet but is much more common at altitudes above 10,000 feet. Characterized by mental confusion, hallucinations and drunkenlike walking, high-altitude cerebral edema often develops within 36 hours after arrival at high altitude.

High-altitude pulmonary edema routinely occurs above 9,000 feet, although it afflicts some people at lower altitudes. The symptoms—including shortness of breath, severe cough, blood-tinged sputum, headache, lethargy and mild fever—usually develop after 36 or 72 hours at altitude.

Acute mountain sickness affects 15 to 17 percent of people who climb to 8,000 feet or higher too rapidly. It is characterized by headache, fatigue, shortness of breath, disturbed sleep and, sometimes, nausea. The illness rarely requires any treatment other than descent.

dividual is stricken with altitude illness.

The dual control of breathing causes an interesting phenomenon—an irregular fluctuation as control shifts between centers in the midbrain responsive to carbon dioxide and blood pH and the carotid bodies, small collections of cells in the neck that are sensitive to oxygen. The result is an erratic pattern called periodic, or Cheyne-Stokes, breathing that is common above 9,000 feet and universal at higher altitudes. Typically a period of rapid, increasingly deep breathing is followed by shallower breathing until breathing stops completely for an alarming time (between eight to 10 seconds), before the cycle repeats. Periodic breathing is more pronounced during sleep, and so average oxygenation falls when one is asleep.

Another early response to hypoxia is an increase in heart rate and stroke volume, which propels more oxygenated blood throughout the body. Concurrently, there is a temporary shift of fluid from blood into tissues, concentrating hemoglobin and thus enabling the heart to deliver more oxygen in each stroke. The increased cardiac output subsides in a week or so, but the increased ventilation persists for the most part throughout the stay at altitude.

Production of red blood cells is stimulated by an immediate increase in erythropoietin—a hormone that acts on bone marrow. Other primary or secondary changes occur in many hormonal systems and in the activity of the sympathetic nervous system. There is little doubt that hypoxia also disturbs electrolyte and water balance. It leads to vasoconstriction, water retention and changes in the permeability of cell membranes as well as disturbed kidney function.

While we understand many of the changes the body goes through in its response to falling concentrations of oxygen, puzzles remain—particularly on the cellular level. The precise pathophysiology of hypoxia has yet to be fully explained.

One theory holds that hypoxia causes a reversible breakdown in the function of an energy-intensive ion channel called the sodium pump. This pump maintains normal levels of sodium and potassium within each cell. Because the sodium pump uses as much as 20 percent of the body's total oxygen uptake, it is not surprising that it may falter when oxygen is in short supply.

According to this theory, when the system falters, sodium accumulates within the cell and potassium leaks out, disturbing water balance and causing edema. The cells in which the pump fails most dramatically are the ones most affected by hypoxia. This failure in turn affects which form of altitude sickness develops. Recent studies also suggest that calcium channels are altered by hypoxia, and it may be that the failure of still other pump systems contributes to mountain sickness.

For reasons that remain unclear, hypoxia also seems to increase the contractility of the small arterioles in the lungs. This resistance causes an increase in pulmonary artery pressure. Consequently, the vessels become distended, stretching the endothelial lining and bringing about the release of biologically active substances, called eicosanoids, or kinins. Some eicosanoids increase vascular leakiness and platelet agglutination, or clumping, whereas others decrease these effects. The strength of the response determines whether or not lung edema results.

Increased pressure in the pulmonary arteries is also thought to stretch and even to rupture the tight junctions between endothelial cells. Together with the action of eicosanoids, the widened junctions permit leakage of plasma and red blood cells into the interstitial and alveolar spaces. This chain of events could explain why high-altitude pulmonary edema is so different from the edema caused by toxic substances, heart failure or injury—in these cases, the membranes of capillaries and alveoli are damaged, but pulmonary artery pressure is not increased.

Unlike the cellular changes accompanying high-altitude pulmonary edema, the mechanisms underlying cerebral edema are elusive. Brain scans of patients have shown both generalized and local swelling. But these findings have been unsuccessful at linking pathology to symptoms. Autopsies have shown generalized edema, scattered small hemorrhages and large clots.

High-altitude residents who acquire chronic mountain sickness have a blunted ventilatory response to lack of oxygen and are thus more hypoxic. In addition, erythropoietin seems to work overtime, creating a proliferation of red blood cells. Together with the exaggerated hypoxia, the greater red blood cell mass increases blood viscosity and can lead to congestive heart failure.

If sea-level residents can be so seriously affected by going to high altitudes, how is it that some mountaineers can climb and stay as high as

20,000 feet and even briefly reach the summit of Mount Everest (at 29,029 feet), where atmospheric pressure is less than a third that of sea level? How is it possible for Tibetans to live on the high plateau of central Asia or the Quechuas to survive in the high Andes? The answer is acclimatization: integrated changes in bodily functions that restore tissue oxygenation toward sea-level values.

Studies of acclimatization in mountain residents or climbers are difficult. Many factors, including cold, exertion, inadequate food and water as well as rigorous living conditions, can confound results. The best way to isolate the effects of hypoxia is to study subjects in a decompression chamber. (This situation, however, has its own stresses: confinement, boredom and lack of exercise can cloud the observations.)

In 1946 Richard L. Riley, then at Bellevue Hospital, and I conducted a study called Operation Everest during which four men spent 35 days in a small steel chamber that was gradually decompressed to a pressure equivalent to that on the summit of Everest. We were able to prove that men could survive and even work lightly there after 10 days of living in an environment equivalent to being at 25,000 feet. (The experiment took place seven years before Everest was climbed using oxygen tanks and 32 years before the first ascent using no supplemental oxygen.)

In 1985 a team of scientists repeated this experiment in Operation Everest II. Eight men lived for 40 days in the chamber, and six repeatedly "went to the summit" after 10 days of residence at conditions equivalent to 24,000 and 25,000 feet. From these and a few other mountain studies have come a better, though incomplete, understanding of acclimatization. Many of us are confident that the lessons learned at high altitude can be important in the diagnosis and care of patients who are hypoxic from illness or injury at sea level.

As I described above, the body initially responds to altitude by struggling to sustain a near-normal oxygen supply to the cells. These emergency responses—increased breathing and heart rate—protect the body while more lasting changes take place. Increased breathing persists as the most powerful response. In addition, new hemoglobin is formed, and the number of red blood cells increases. Normally, between 20 and 30 percent of the body's capillaries are inactive, and at altitude some of these tiny blood vessels are recruited for more efficient dispersal of blood to tissues. Within the cells, enzyme changes enhance anaerobic metabolism.

Although the body can partly compensate for lowered levels of oxygen, many activities remain impaired. Researchers have shown that every 1,000-foot increase in altitude reduces maximum work capacity by 3 percent. Long residence cannot fully restore work capacity toward that at sea level.

Descendants of people living for many generations at high altitudes show more permanent changes that are probably genetic. Some natives have larger lungs within a barrel chest; others have a substantially increased amount of hemoglobin. More or larger or differently placed mitochondria (the tiny factories that power the cells) are found in populations such as the Quechuas of Chile.

Despite centuries of residence, however, no humans have adapted permanently to altitudes higher than 17,000 feet. Sea-level dwellers can live for only a few months at this height before deterioration outstrips acclimatization. Over time, people continue to lose weight, their motivation decreases and all bodily functions decline.

Studies of animals offer tantalizing glimpses into the many strategies by which oxygen can be transported and used. Diving seals and whales, for instance, have very large spleens, which serve as a reservoir from which they can draw oxygenated blood. Like humans, their major accommodation is the diversion of blood flow from less to more essential organs, although, unlike us, they shut down certain functions entirely. Other animals, such as yaks and llamas, use different forms of hemoglobin to obtain and hold more oxygen.

For people, acclimatization is best ensured by slow ascent. Indeed, the impatient are likely to become patients. It is a good rule to ascend no more than 2,000 feet a day when above 7,000 feet and to climb at a rate well tolerated by the most vulnerable member of the party. If symptoms become prominent, take a day of rest or even descend a few hundred feet at night. The altitude at which one sleeps is more important than the altitude reached during the day. It is also necessary to drink more water at altitude than at sea level to compensate for the fluid lost through overbreathing. Avoiding strenuous exercise for the first day or two is helpful. Taking more salt than usual tends to cause water retention, perhaps enough to trigger altitude sickness.

Those who wish a quick fix may take certain prescription drugs. Acetazolamide inhibits the secretory activity of an enzyme called carbonic anhydrase and thus enables deeper or faster breathing without excessive loss of carbon dioxide. The drug also inhibits the action of antidiuretic hormones. Dexamethasone, a glucocorticoid, is helpful partly because of the euphoria it causes, although it can sometimes have serious side effects, including psychotic episodes. Nifedipine, a calcium channel blocker and general vasodilator, lowers pulmonary artery pressure. Its long-acting form may protect those climbers who are unusually susceptible to high-altitude pulmonary edema.

As for other treatments, the mild forms of illness usually pass in a few days. If the more serious pulmonary edema or cerebral edema is suspected, descent is the best and most important remedy. And getting down only a few thousand feet brings rapid relief. If not, some other illness should be suspected. Breathing extra oxygen is helpful, but on a mountain where the supply is limited, and in serious cases of edema, oxygen should not delay descent—getting down is always the best approach. When descent is impossible, a strong diuretic or morphine can be used. A recent and promising treatment is to place the patient in a bag that can be inflated to a pressure equivalent to a much lower altitude.

As the relatively recent recognition of high-altitude pulmonary edema suggests, we have much to learn about altitude sickness. Further insights will make going to the mountains safer and more enjoyable, if visitors follow simple rules. A fuller understanding of hypoxia could also hold relevance for the treatment of conditions not related to altitude, including emphysema, asthma and pneumonia.

But as studies of acclimatization and eventual deterioration show, it is clear that certain physiological limits do exist. Perhaps it is fitting that the highest peaks remain in some way beyond our grasp, exerting their powerful pull and forever challenging us.

FURTHER READING

HIGH ALTITUDE PHYSIOLOGY. Edited by John B. West. Van Nostrand Reinhold, 1981.

THE MONTGOLFIER BROTHERS AND THE INVENTION OF AVIATION, 1783–1784. Charles S. Gillespie. Princeton University Press, 1983.

GOING HIGHER: THE STORY OF MAN AND ALTITUDE. Charles S. Houston. Little, Brown and Company, 1987.

HIGH ALTITUDE MEDICINE AND PHYSIOLOGY. Michael P. Ward, James S. Milledge and John B. West. University of Pennsylvania Press, 1989.

Histones as Regulators of Genes

Histones were once dismissed as little more than packing material for nuclear DNA. In fact, these proteins can both repress and facilitate activation of many genes

by Michael Grunstein

As recently as five years ago, many investigators assumed that small proteins called histones were irrelevant to the regulation of genes. Histones combine in the cell nucleus with long strands of DNA, which contain the genes, to form chromatin, the material that makes up the chromosomes.

The proteins were discounted mainly because a number of researchers who studied how genes function had concluded that histones probably serve as nothing more than cellular packing material. The proteins seemed, in short, to do little more than form positively charged "spools" around which negatively charged strands of DNA were wound so they could fit inside the tiny cell nucleus.

Yet study after study now indicates that histones are vital participants in gene regulation. At least one type helps to activate genes: it helps to initiate the copying, or transcription, of information stored as DNA into molecules of messenger RNA. (These transcripts are the templates from which individual proteins are synthesized.) Certain histones can also repress transcription. Today molecular biologists are beginning to realize that if they want fully to elucidate how genes are controlled, they must consider the roles of histones.

Investigators are preoccupied with the problem of how genes are turned on and off, in part because such knowledge will help explain how embryonic development unfolds in multicellular organisms. Although virtually every cell in a developing organism has the same complement of genes, some differentiate to become neurons, whereas others become, say, blood or liver cells. The end states vary largely because different genes are switched on or off at appropriate times, yielding the distinct mixtures of enzymes and other proteins that give a cell its unique properties.

Insight into gene activation and repression should also help explain how the controls on these processes sometimes go awry, leading to disease. For instance, cancer can arise when genes whose protein products facilitate the replication of cells become overly active or when genes that suppress replication become abnormally still.

MICHAEL GRUNSTEIN, who was born in Romania, is professor of molecular biology at the University of California, Los Angeles. He earned a B.S. from McGill University in 1967 and a doctorate in molecular biology from the University of Edinburgh in 1971. Grunstein then completed postdoctoral fellowships in the departments of medicine and biochemistry at Stanford University before moving to U.C.L.A. in 1975.

The regulatory functions of histones eluded discovery until recently because scientists have been able to induce transcription by exposing naked DNA (unbound by histones) to cellular extracts containing selected regulatory proteins. The success of these "cell-free" experiments led many biologists to conclude that histones were uninvolved in gene regulation in living organisms.

Although that conclusion was incorrect, these studies and others generated profound insights into how genes are activated. That research, together with structural analyses of chromatin, also paved the way for much of the work that ultimately uncovered the importance of histones.

Critical early discoveries, for instance, established that the genes of eukaryotic (nucleated) cells include at least three specialized segments. Among these, naturally, is the coding region, which specifies the amino acid sequence of a protein. The others are distinct regions that influence whether the coding segment is copied into messenger RNA and how many RNA molecules are made.

For a gene to be switched on, a particular set of proteins must assemble on a regulatory region often called the proximal promoter. First, one protein binds to a part of that promoter known as the *TATA* box. Then other proteins join with the first, giving rise to a combination of protein and DNA—the preinitiation complex. The *TATA* box is so named because it includes some approximation of the nucleotide sequence *TATAAATA*. Nucleotides, the building blocks of DNA, are distinguished from one another by one of four chemical bases: thymine (*T*), adenine (*A*), guanine (*G*) or cytosine (*C*).

Once the preinitiation complex has formed, it positions one of its members—the enzyme RNA polymerase—at a spot within the proximal promoter called the transcription-initiation site. The accurately positioned polymerase then travels up to and down the length of the coding region (perhaps like a train on a track), synthesizing messenger RNA. Because recognition of DNA by the proteins of the preinitiation complex will, under the right conditions, generate a low, or basal, level of transcription in the test tube, the proteins are sometimes called basal factors.

Binding by other proteins, known as activators, to a second regulatory site—the upstream activator sequence—gives rise to maximal production of the messenger. (In many eukaryotic organisms, upstream activator sequences are also referred to as enhancers.)

Even before many of these details emerged, research into the chemistry and structure of chromatin had established that histones come in five basic forms, designated H1, H2A, H2B, H3

HISTONE PROTEINS H3 (*blue*) and H4 (*green*) exhibit an ability to bend DNA (*golden brown ring around the histones*) in this thin, front-to-back slice through a three-dimensional map of a nucleosome. Nucleosomes are a prominent structural feature of the chromosomes in all nucleated cells. Histones have only recently been shown to play a role in regulating the activities of the genes that are stored in chromosomes.

and H4. X-ray crystallography and other techniques subsequently showed that the "spools" around which DNA is wound are octamers; they consist of eight molecules—two of histone H4, combined with two of histones H3, H2A and H2B. A stretch of DNA some 146 nucleotides long wraps almost twice around the octamer, or histone core particle, yielding a unit known as a nucleosome [see "The Nucleosome," by Roger D. Kornberg and Aaron Klug; SCIENTIFIC AMERICAN, February 1981].

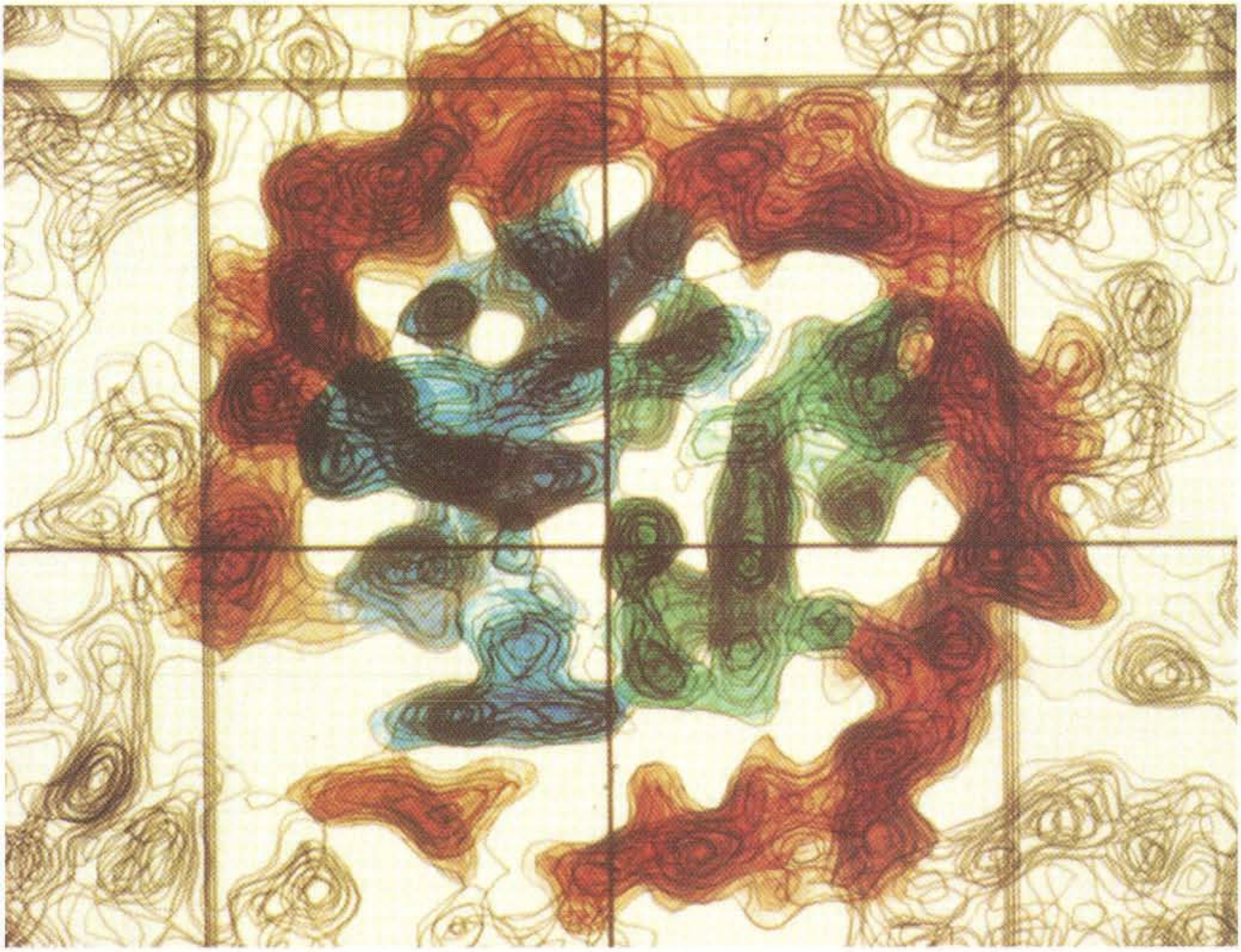
In most eukaryotic organisms, the DNA is further locked in place by a single molecule of histone H1. This protein binds to the outer surface of the DNA that encircles the octamer as well as to both stretches of the flanking, or "linker," DNA that connects one nucleosome to another. Histone H1 is not essential to the formation of nucleosomes, however. For instance, brewer's, or baker's, yeast (*Saccharomyces cerevisiae*), the single-celled eukaryote that has provided many clues to how genes are regulated, apparently produces very little (if any) of it; yet nucleosomes abound in the yeast's chromosomes.

On the other hand, H1 seems to promote the compacting of DNA in cells of multicellular organisms. In those cells, much of the chromatin coils into fibers that have a diameter of about 30 nanometers (billionths of a meter); these coils include about six nucleosomes per turn. In contrast, most of the chromatin in brewer's yeast seems to remain in a basically extended conformation. It resembles a string (DNA) of beads (nucleosomes) and, for the most part, is about 10 nanometers in diameter—the measure of a single nucleosome.

By the mid-1980s several investigators, aware of these structural and regulatory findings, had begun to test seriously a controversial notion: *TATA* boxes, upstream activator elements and the specialized proteins that bind to them (basal factors and activator proteins) might not be the only participants in gene regulation; histones might play a role as well. Earlier, in the 1960s, a few studies had hinted that histones might repress transcription. But the data were weak.

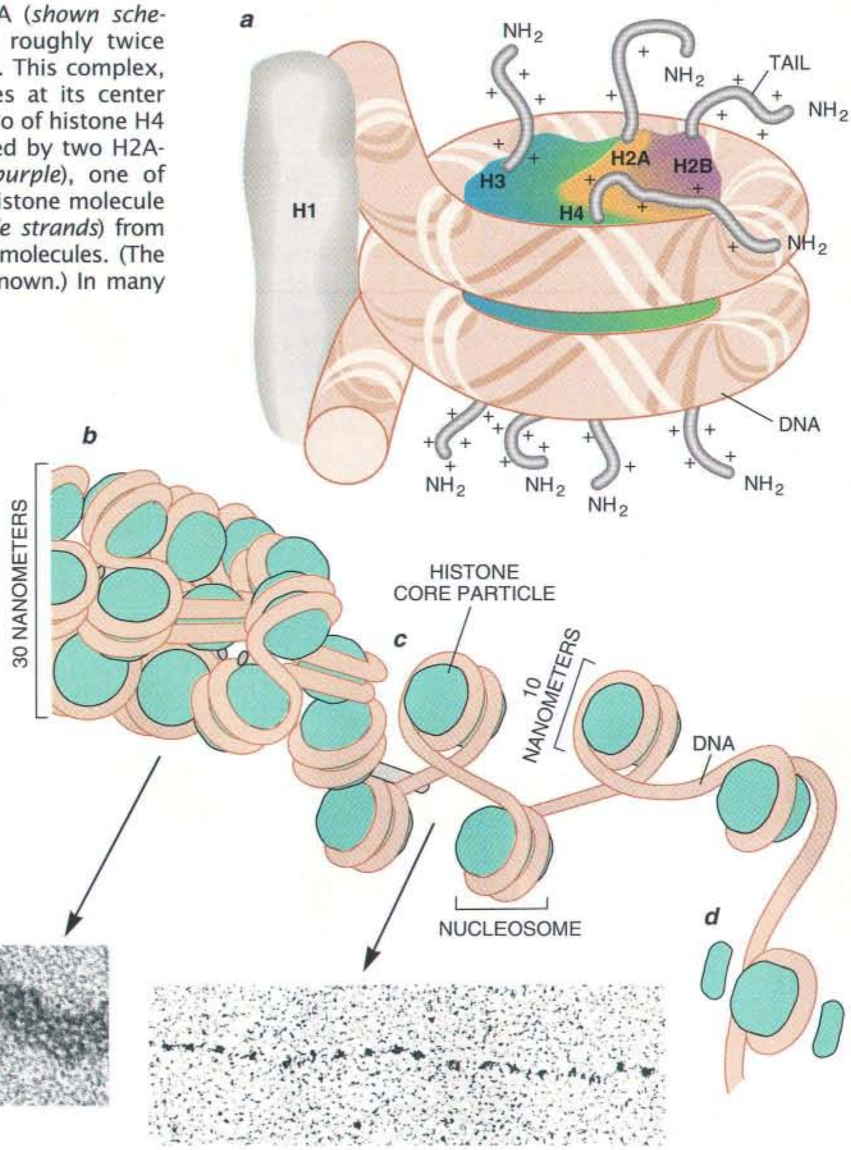
Some 20 years later investigators found themselves in a position to conduct more definitive studies. Researchers were also intrigued by an evolutionary discovery. Analyses of the amino acid sequences of the histones had established that the proteins of the histone core differ little from one species to another. Indeed, the histones in peas are much like those in cows. When the structure of a molecule varies minimally among species, the lack of diversity implies that the amino acid sequence is very important to a cell's functioning. If folding of DNA were the histones' only job, almost any sequences rich in positively charged amino acids might be able to bend the negatively charged DNA; strict conservation of precise amino acid sequences over millions of years would be unnecessary.

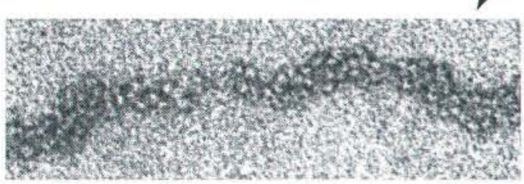
Part of the impressive support for the notion that histones play a role in controlling genes came from a series of cell-free studies. In 1986 Joseph A. Knezetic and Donal S. Luse of the University of Cincinnati College of Medicine combined histone proteins with DNA molecules containing genes of a human adenovirus. When the scientists then intro-



The Structure of the Nucleosome

A nucleosome (*a*) forms when histones combine with one another, causing DNA (*shown schematically as a tube*) to wrap roughly twice around the resulting histone complex. This complex, or nucleosomal core particle, includes at its center two copies of histone H3 (*blue*) and two of histone H4 (*green*). This H3-H4 tetramer is flanked by two H2A-H2B pairs, or dimers (*yellow and purple*), one of which is not visible. An end of each histone molecule is thought to protrude like a tail (*pale strands*) from the core, ready to interact with other molecules. (The exact conformation of the tail is not known.) In many (but perhaps not all) organisms, histone H1 (*gray structure at left*) helps to anchor DNA to the core. Together nucleosomes and the histone-free DNA that joins them to each other form chromatin (*b* through *d*), the material from which chromosomes are fashioned. By promoting the tight packing of nucleosomes, histone H1 helps chromatin to coil into 30-nanometer fibers (*drawing and micrograph in b*). Genes housed in the DNA within these 30-nanometer-thick fibers are thought to be silent. Unfolding of the chromatin into a 10-nanometer conformation (*c*) is more conducive to gene activation. For genes to actually be turned on, DNA must partially unwind from nucleosomes at DNA sequences involved in regulating genes (*d*).





duced fractions of human cells that included basal factors (among them RNA polymerase), the newly formed nucleosomes blocked the factors from initiating transcription.

Then Roger D. Kornberg's group at Stanford University showed that selective inclusion of *TATA* boxes in nucleosomes would prevent RNA polymerase from initiating transcription. Kornberg's team and, independently, that of Donald D. Brown of the Carnegie Institution of Washington made another important discovery as well. When nucleosomes were induced to form within the coding region, they did not block transcription (although later studies were to show that nucleosomes could slow the passage of RNA polymerase).

Collectively, these results implied that histones could potentially help repress genes within whole cells. They might act by preventing basal factors from locking onto DNA rather than by completely halting transcription within the coding region. Conversely, it seemed likely that if genes were to be activated, histone core particles would have to be disengaged from the *TATA* box.

Other cell-free tests implicated activator proteins—the molecules that bind to upstream activator sequences—as having a possible disruptive influence on nucleosomes. As early as 1984, Beverly M. Emerson and Gary Felsenfeld of the National Institutes of Health had shown that proteins that bind to regulatory sequences in DNA can prevent nucleosomes from forming at those same sites. Later the laboratories of Luse, Robert G. Roeder of the Rockefeller University and others independently extended the finding. They showed that histones and basal factors compete for access to the *TATA* box.

These teams varied the order in which they exposed DNA to histones and basal factors. Transcription occurred only if the factors were given the opportunity to assemble on *TATA* elements before histones were added. Histones won, however, if the competitors were added to DNA simultaneously.

In the course of related experiments, Jerry L. Workman, Roeder and their colleagues at Rockefeller made a curious discovery. Their data suggested that if

DNA was simultaneously exposed to histones and cellular extracts containing both basal factors and activator proteins, the odds tilted in favor of the basal factors. They bound to *TATA* boxes in spite of the presence of histones, thereby preventing the *TATA* boxes from being incorporated into nucleosomes. James T. Kadonaga and his associates at the University of California at San Diego have similarly shown that activators will prevent repression of transcription by histone H1 in vitro.

Such results raised the interesting possibility that activator proteins might directly or indirectly help to loosen the ties between histones and DNA in cells. For instance, an activator might alter some basal factor in the preinitiation complex, enabling that protein to disrupt nucleosomes and gain access to the *TATA* element.

Studies of cell extracts are extremely useful. But the results do not necessarily accurately reflect how cells operate within intact organisms. Consequently, in parallel and sometimes in response to the cell-free work, several laboratories, including mine at the University of California at Los Angeles, focused on the role of histones in living cells. We asked such questions as, are *TATA* boxes routinely incorporated in nucleosomes? Do histone core particles repress transcription when *TATA* elements are so engaged? How might the bonds between histone octamers and DNA be weakened in established nucleosomes?

Dennis E. Lohr of Arizona State University and Wolfram Horz of the University of Munich and their colleagues gathered some of the strongest evidence that nucleosomes often form at *TATA* boxes. They demonstrated that many genes in brewer's yeast consistently display nucleosomes at the *TATA* element. The genes include, among others, those specifying the enzymes galactokinase (*GAL1*) and acid phosphatase (*PHO5*).

The investigators revealed, too, that activation of such genes was accompanied by unwinding of DNA from nucleosomes associated with *TATA* boxes. The researchers knew a degree of dissociation was occurring in active genes because nucleases (enzymes that cleave DNA) were able to cut the regions containing the *TATA* element more easily. Such cleavage can occur only when the DNA unbends enough to become accessible to the nucleases.

Yet the fundamental question of cause and effect remained. Did changes in the structure of nucleosomes enable transcription to occur, or were these changes merely a consequence of the passage of RNA polymerase along the gene?

In the late 1980s my laboratory attacked the problem by manipulating the histone genes of intact brewer's yeast. Like many biologists, we favored this simple, unicellular organism partly because the brewing industry and other biologists had already deciphered much of its genetic makeup. Equally important, tools had been developed for splicing and combining yeast genetic elements for experimental purposes.

We reasoned that if we could shut down histone synthesis, we could prevent the *TATA* boxes of many normally repressed genes from becoming trapped in nucleosomes. If messenger RNAs for those genes were then produced, we would have shown that removal of histones enables transcription to occur not just in the test tube but also in living cells. The finding would also indicate that histones can repress genes in living cells.

My graduate students Min Han and Ung-Jin Kim halted histone synthesis by adapting a technique devised by Mark Johnston of the Washington University School of Medicine. They fused regulatory elements from the yeast galactokinase gene to the coding regions of the genes that give rise to histones. Then they substituted these genes for the normal versions in yeast cells. When cells that are so altered are grown in a glucose-enriched medium, the regulatory galactokinase sequence prevents the associated genes from being transcribed. Hence, to halt histone synthesis, Han and Kim simply had to expose the genetically altered yeast to glucose.

At the beginning of treatment, the cells were all at different stages of the yeast life cycle. Unlike many mammalian cells (such as neurons), brewer's yeast cells continually grow and divide (by budding) unless environmental conditions threaten survival of the new cells. As the organism's chromosomes are duplicated during cell division, the two sets are segregated. One is delivered to an emerging bud that pinches off to become an independent cell.

After histone synthesis was halted, the genetically altered yeast cells were able to reproduce their DNA one time without difficulty. But the resulting DNA carried approximately half its usual number of nucleosomes. What is more, nuclease-digestion analyses indicated that most of the remaining nucleosomes were no longer in their usual positions. Instead they were much more randomly placed. Further, our findings argued that *TATA* boxes that were normally found within nucleosomes were now freed. In that way, the chromosomes resembled naked DNA and thus fit our experimental needs quite well.

Han and Kim proceeded by measuring the amount of messenger RNA produced by many different genes in the nucleosome-depleted yeast. In 1988 they learned that the only genes switched on in response to the depletion of nucleosomes were those of the inducible type. Such genes remain silent unless a cell is exposed to a particular stimulus, such as a change in the level of a specific sugar or amino acid. Conversely, so-called housekeeping genes, whose products are constantly needed for cellular self-maintenance, did not become unusually active.

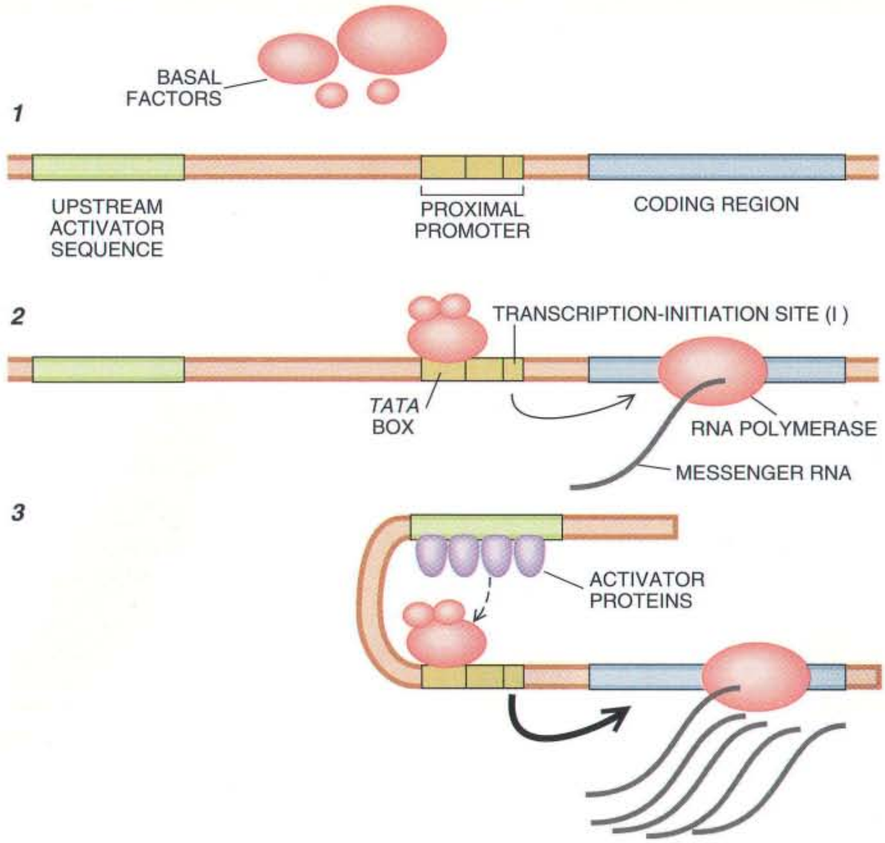
This last result is not surprising if the proposal that nucleosomes repress transcription is correct. Genes that are continuously in use would be expected to have lacked nucleosomes at their *TATA* boxes even before the experiments started. Hence, procedures designed to remove histones from DNA would not affect them.

Closer examination of the inducible genes yielded more direct support for the repression hypothesis. By means of various genetic manipulations, Han and a postdoctoral fellow, Linda K. Durrin, found that for many such genes, elimination of nucleosomes at *TATA* boxes led to synthesis of messenger RNA. Such synthesis suggested that basal factors had assembled on the *TATA* boxes. Transcription occurred even in the absence of upstream activator sequences, which indicated that removal of *TATA*-associated nucleosomes is sufficient to initiate a basal level of transcription. For some genes, however, reinstatement of the upstream activator sequence was needed to increase transcription to the maximal rate.

My colleagues and I interpret these results to mean that, by forming nucleosomes, histones do indeed repress transcription in living cells. On the basis of these data and those of in vitro studies, we have proposed a revised, two-stage model of how genes are turned on.

In the first, histone-dependent, activation stage, activator proteins at the upstream activator sequence directly or indirectly cause the histone core particle to dissociate from the *TATA* box. This displacement enables basal factors to assemble on the *TATA* box and form a preinitiation complex, thereby generating a basal level of transcription.

In the second, histone-independent, stage, activators stimulate the preinitiation complex to maximize production of messenger RNA. We think we can now explain why molecular biologists



RECENT MODEL OF GENE ACTIVATION focuses on naked, or histone-free, DNA. Genes include a coding region (*blue band at far right in 1*), which specifies the amino acid sequence of a protein, and two major regulatory regions—the proximal promoter (*yellow band*) and the upstream activator sequence (*green band*)—which determine whether the encoded protein is made. The model holds that if a gene is to be turned on (*2*), proteins called basal factors (*red*) must assemble on the *TATA* box within the promoter (*leftmost yellow segment*). Such an assembly positions one basal factor—the enzyme RNA polymerase—on the transcription-initiation, or I, site (*rightmost yellow segment*), enabling the enzyme to copy, or transcribe, the information in the coding region into messenger RNA (the template from which an encoded protein is synthesized). Binding of so-called activator proteins (*purple in 3*) to the upstream activator sequence yields further stimulation of the basal complex (*broken arrow*) and maximizes transcription (*dark arrow*).

can sometimes achieve activation in vitro in the absence of histones: they may be replicating only the final events of the activation process.

In 1988 Fred Winston of Harvard Medical School and his co-workers amassed other cellular evidence that histones participate in gene regulation. Winston's group built its technique around the discovery by Gerald R. Fink and his associates at the Whitehead Institute that in some mutant yeast cells a transposable segment of DNA called TY inserts itself near the *TATA* box of the gene *HIS4*. (*HIS4* does not specify a histone; it encodes an enzyme involved in use of the amino acid histidine.) The insertion causes transcription to begin within the TY element rather than at the start of the *HIS4* gene. This error results in production of a nonsensical transcript that cannot direct synthesis of the His4 protein.

Winston's team demonstrated that altering the number of copies of the genes for histones H2A and H2B in yeast overcame the repressive effect of TY, leading to accurate transcription of the *HIS4* gene. This reversal of *HIS4* repression signified that histones were probably more important in gene regulation than was once thought.

The accumulating evidence strongly argues that one cannot explain gene regulation merely by positing interactions between DNA and nonhistone proteins. Removal of histone-controlled repression is a necessary first step in the turning on of a silent eukaryotic gene. If this conclusion is correct, it becomes important to explain how the ties between the histone core and the *TATA* box are relaxed.

My colleagues and I believe such relaxation probably results from specific binding between activator proteins or an intermediary and the "tail" segment of one of the histones: H4. Each histone in the octamer spool is divided into two major domains. The core domain, which includes the carboxy (COOH) terminus of the protein, twists into tight hydrophobic (water-repelling) helices. These helices mesh to form the bulk of the histone octamer. In contrast, the tail domain, which includes the amino (NH_2) terminus, is hydrophilic (water-attracting). It remains uncoiled, perhaps resembling a floppy string.

We were drawn to examine a regulatory role for histone tails, in part because the laboratories of Harold Weintraub, now at the Fred Hutchinson Cancer Research Center in Seattle, and of James P. Whitlock and Robert T. Simpson of the NIH had shown that these "strings" have little to do with the assembly and stability of nucleosomes. They do, however, extend from the helices. Compared with the core domains, they are thus potentially more available to interact with molecules in the local environment of the chromatin.

We were also interested in the tails because of other intriguing reports. As early as 1977, Vincent G. Allfrey and his colleagues at Rockefeller had observed that transcription was often accompanied by the addition of acetyl groups (CH_3CO) to the histone tails. Such additions would be expected to neutralize the positively charged tails, which in turn could potentially disrupt the interaction of the tails with negatively charged DNA. Whether acetylation precedes or follows transcription is still not clear. But the finding did suggest that the tails might be important for freeing *TATA* boxes from nucleosomes.

In 1991 Durrin uncovered direct support of that likelihood. By deleting amino acids 4 through 23 in the tail of histone H4, she strongly inhibited transcription of several normally inducible genes in yeast, including those involved in metabolism of the sugar galactose. The inhibition intimated that all or part of the eliminated segment normally interacts with the transcription machinery to make transcription possible.

Durrin's finding has led my colleagues and me to amplify the histone-dependent activation stage in our model of gene regulation. We propose that activator proteins, or proteins they influence, displace histones from the *TATA* box by specifically binding to some part of the region between amino acids 4 and 23 on the tail of histone H4.

We can envision a way in which such binding might facilitate transcription. Our ideas are inspired in part by an understanding of how nucleosomes form

and how they behave when they are unfolded. Part of this awareness was originally contributed by the late Abraham Worcel, then at Princeton University. He suggested that the first step in nucleosome construction is the assembly of two histone H3 molecules and two histone H4 molecules into a foursome, or tetramer. DNA then wraps around the tetramer, making contact with each of the four molecules. (Felsenfeld and Alan P. Wolffe of the NIH and Kensal E. Van Holde of Oregon State University have ascertained that the tetramer is stable and behaves like a complete nucleosome in many ways.)

There is also evidence that this partial structure directs the subsequent addition of two histone H2A-H2B pairs, or dimers, to the tetramer. This addition in turn promotes further coiling of DNA around the core particle. Bradford B. Baer and Daniela Rhodes of the Medical Research Laboratory of Molecular Biology in England have found as well that nucleosomes depleted of their H2A-H2B dimers are less effective at blocking transcription than are intact nucleosomes. Moreover, work by Vaughn Jackson and Roger Chalkley and their colleagues at the University of Iowa indicates that histones H2A and H2B may dissociate easily from the tetramer and transfer from one nucleosome to another.

Such insights imply that when activator proteins or their accessories bind to a specific spot on the tail of histone H4, the binding may induce a conformational change in the histone. This change may lead to temporary ejection of histones H2A and H2B from the nucleosome. Displacement of the dimers could in turn cause some of the DNA in the nucleosome to uncoil and thus become accessible to basal factors. After passage of the RNA polymerase, the H2A-H2B dimers may reassemble on H3-H4 tetramers to re-form nucleosomes behind the polymerase.

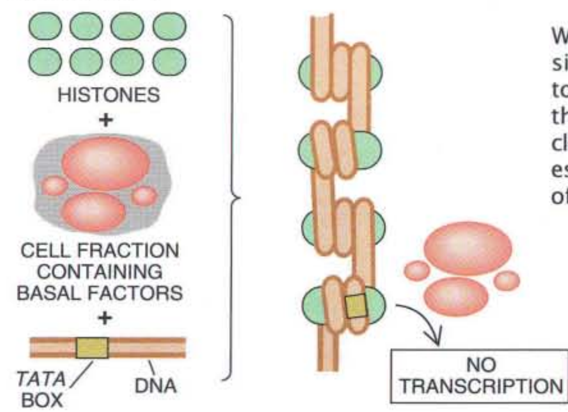
Strangely enough, as we examined different segments of the histone H4 tail, we realized that its parts had a number of different and sometimes conflicting responsibilities. For instance, although certain amino acids between numbers 4 and 23 are needed for transcription of several genes, a part of the tail participates in repression of other genes.

Paul S. Kayne, a graduate student in my laboratory, has identified an important inhibitory function of the histone H4 tail. It helps to repress what are called silent-mating loci. There are two such genetic regions in each brewer's yeast cell. They must remain perpetually quiet, or else the cells in which they are

Critical Test-Tube Experiments

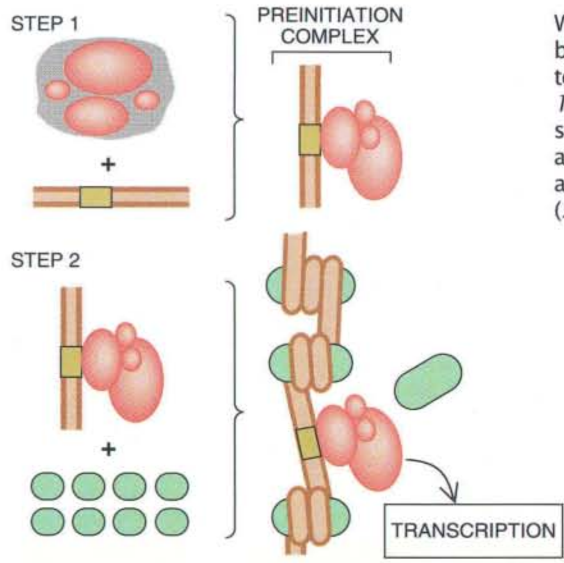
In the late 1980s studies of extracts from human cells demonstrated that histones and basal factors compete for access to *TATA* boxes and that histones generally win (*experiment 1*), except under special circumstances (*experiments 2 and 3*). These findings suggested that activator proteins or proteins they influence might be the agents that free *TATA* boxes from nucleosomes in intact cells, allowing transcription to occur.

EXPERIMENT 1



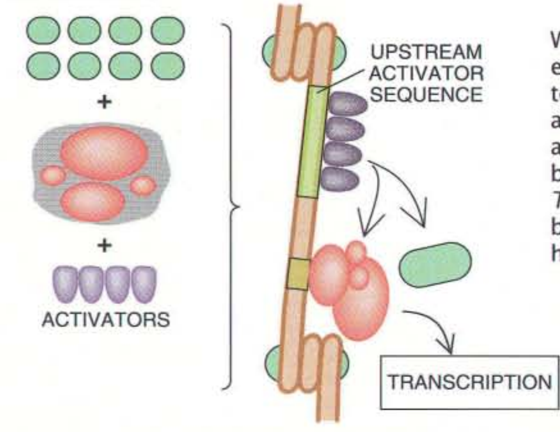
When DNA is exposed simultaneously to histones and basal factors, the histones form nucleosomes at *TATA* boxes and block the access of basal factors.

EXPERIMENT 2



When DNA is exposed to basal factors before histones are introduced (*step 1*), the basal factors assemble on *TATA* boxes and lock out histones that are added subsequently (*step 2*).

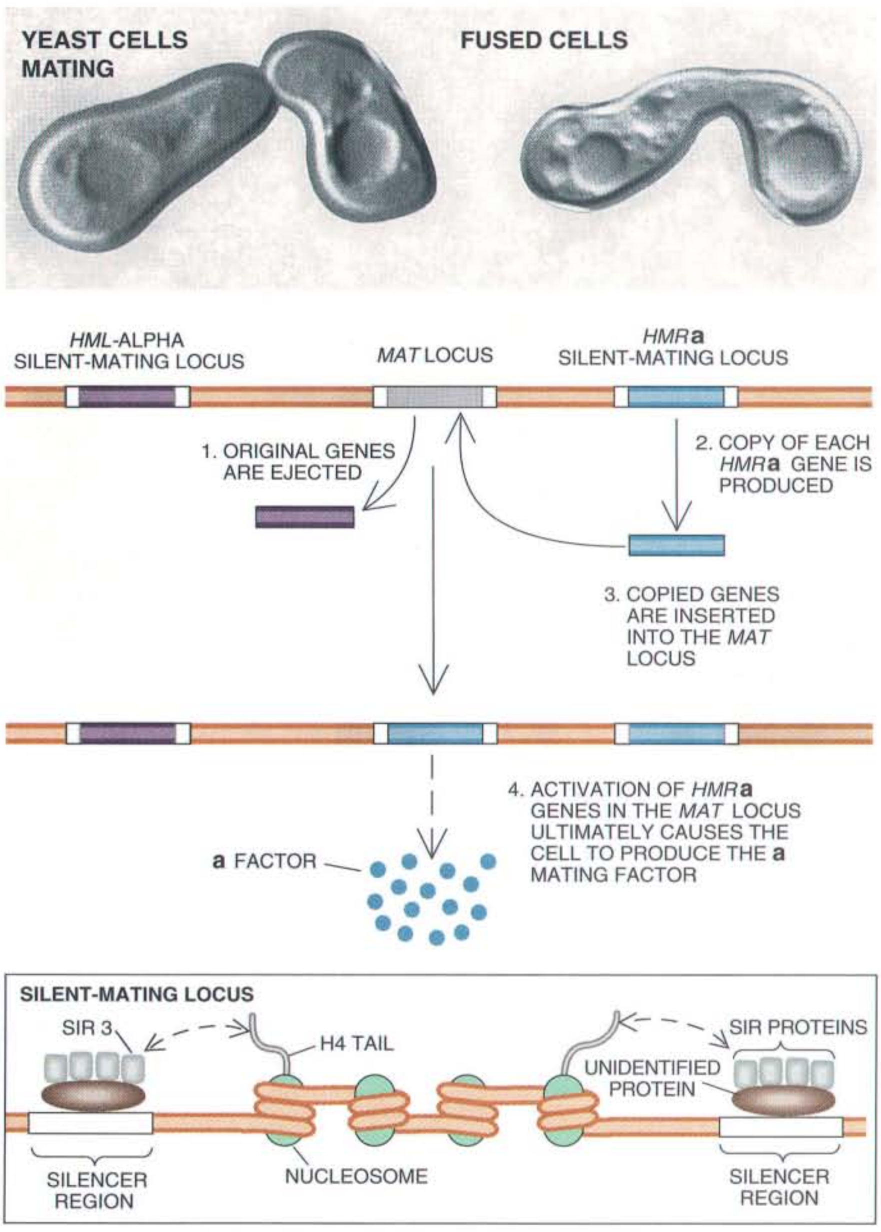
EXPERIMENT 3



When DNA is simultaneously exposed to histones, basal factors and activator proteins, the activators enable the basal factors to bind to *TATA* boxes, and they block access of the histones.

active will become unable to mate—that is, fuse with each other. Yeast cells mate when they are in the haploid state: when they have one full set of chromosomes. Fusion of two haploid individuals yields a single diploid cell, much as the union of a sperm and egg produces a diploid embryo carrying one set of chromosomes from the mother and a matching set from the father.

Kayne showed that deletion of amino acids 15 to 19 from the tail of histone H4 greatly decreased the ability of yeast to mate. The difficulty arose because the genes in the formerly repressed regions became activated. Our laboratory and those of M. Mitchell Smith of the University of Virginia and Jack Szostak of Harvard University found as well that single amino acid substitutions between amino acids 16 and 19 had a similar effect. Hence, each of the amino acids normally present in the region is necessary for repressing the silent-mating loci.



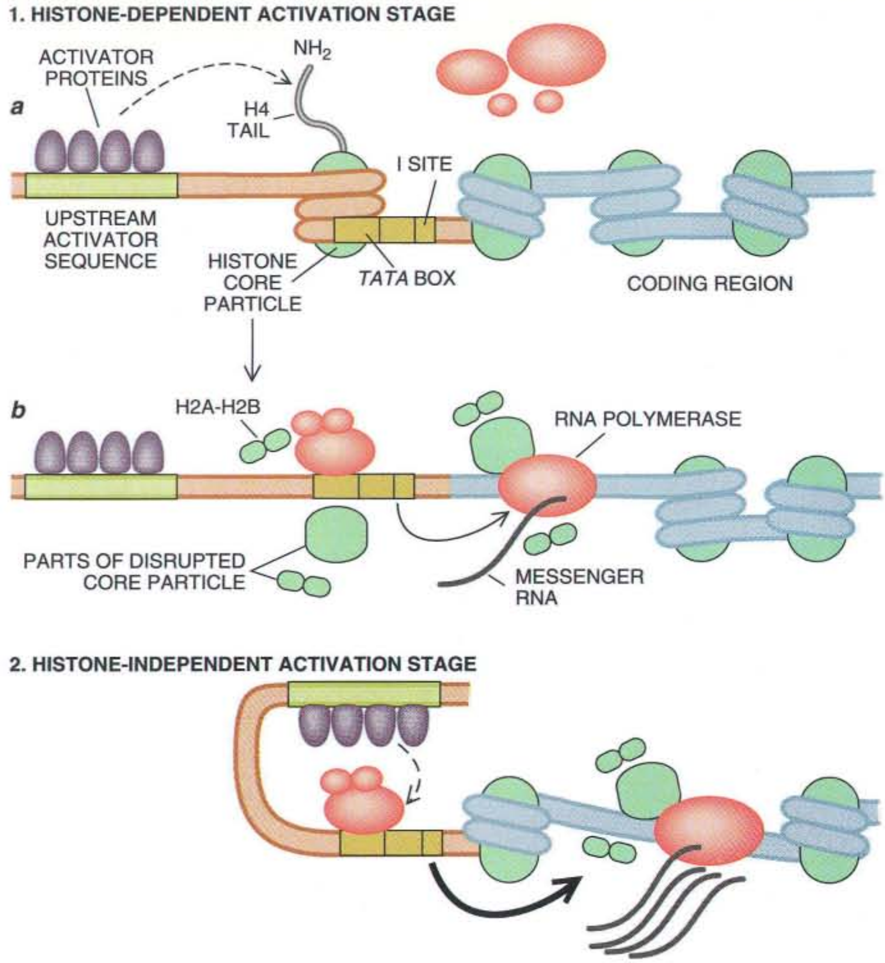
YEAST CELLS have been captured in the process of mating (*left photograph*) and after they have fused (*right photograph*). For fusion to occur, the mating type (or "sex") of one cell must be alpha, and the other, a. Three genetic loci—*HML*-alpha, *MAT* and *HMR*a—on a single chromosome (*top of drawing*) enable yeast cells to switch from one mating type to another. Switching occurs when genes already in the *MAT* locus are discarded (*1*) and replaced with a duplicate copy of genes in either *HML*-alpha or *HMR*a (*2* and *3*), which are called silent-mating loci. The original genes in these silent loci are never activated, but insertion of a duplicate set into the *MAT* locus enables the copies to be transcribed (*4*). Here *HMR*a genes have been copied into the *MAT* locus, causing the cell to become type a. The detail at the bottom shows that a segment of the histone H4 tail helps to keep the silent-mating loci quiet. The tail exerts its effect by interacting somehow (*broken arrows*) with the Sir 3 protein, one of several types of Sir proteins (*gray*) that help silencer DNA sequences at both ends of the loci (*white bands*) to repress genes within the regions.

Lianna M. Johnson in our laboratory recently proved that most of the amino acids between numbers 21 and 29 are additionally required for repression of the same genes. She demonstrated, too, that substitution of either of two amino acids in a protein called Sir 3 can counteract the mating deficit induced by the H4 mutation. The Sir change led to repression of the silent-mating loci and to renewed mating. The fact that a defect in histone H4 is counteracted by a specific defect in Sir 3 argues that histone H4 and Sir 3 normally collaborate to repress the silent-mating loci.

The silent-mating loci reside near the tips, or telomeres, of yeast chromosome III. Daniel E. Gottschling and his colleagues at the University of Chicago have implicated the histone H4 tails in also repressing other genes near yeast telomeres. Gottschling's group found, quite surprisingly, that some of the same H4 mutations that release repression of the silent-mating loci overcome repression of genes that have been artificially inserted close to telomeres. Defects in certain Sir proteins do the same. Hence, it seems that genes residing near telomeres are silenced in much the same way as are the silent-mating loci.

Besides providing concrete new details of how histones function, analyses of mating have generated a tool for further study of histones. A group of genetic defects, collectively called *swi* (switch) mutations, block activation of a yeast gene that specifies an enzyme essential to the establishment of mating type. (Yeast can be type **a** or type alpha.) Ira Herskowitz of the University of California at San Francisco and his team find that certain other mutations (known as *sin* mutations) counteract that blockade. Among these is a mutation in the gene for histone H3, which implies that the altered region of H3 can influence gene activity. Discovery of other *sin* mutations should point to additional gene regulatory segments of histones. Such mutations should also help identify proteins that control genes by influencing histones.

Similarly, Winston's laboratory has determined that the effects of certain so-called *snf* mutations generated by Marian Carlson of Columbia University resemble those of *swi* mutations. The



EXPANDED MODEL OF GENE ACTIVATION holds that transcription in living cells involves histones. The model proposes that for transcription to begin, activator proteins (*purple in 1a*) must interact directly or indirectly (*broken arrow*) with the tail of histone H4 at the *TATA* box (*squiggle emerging from green core particle*). The interaction leads to partial breakup of the core, perhaps by releasing H2A and H2B (*small green fragments in 1b*). The breakup leads to binding of the *TATA* box by basal factors (*red complex*) and to a low level of transcription. Maximal transcription is achieved much as the naked-DNA model of gene activation would suggest (*2*).

proteins produced by the normal versions of the mutated genes may help unravel nucleosomes so that transcription can occur. Elucidation and analysis of *swi*, *snf* and other mutations that can affect chromatin structure are becoming an increasingly important part of research into gene regulation.

Because histone H3 is similar to histone H4 in many ways (including its pattern of acetylation and its role in forming and stabilizing nucleosomes), it would be reasonable to expect that the tail of histone H3 controls genes in much the same way as the tail of H4. For instance, one might predict that the histone H3 tail would help suppress the silent-mating loci. In fact, it does not. Further, Randall K. Mann in my laboratory has reported that the histone H3 tail is required for repression of many of the same genes that need the histone H4 tail in order to become activated.

To explain why histones H3 and H4 have such different functions, researchers will need a better understanding of how these molecules interact with other proteins and with DNA. The cellular roles of histones H2A and H2B also remain to be determined, although it is clear that they do not behave in the same way as histones H3 and H4.

So far much of what is known about how histones regulate genes in living organisms has been gleaned from yeast. In yeast, upstream activator sequences appear to be permanently free of nucleosomes. In human cells, much of the chromatin is usually in the 30-nanometer conformation. Consequently, during development human cells may need to activate some genes whose upstream activator sequences, or enhancers, are incorporated in nucleosomes. How might they manage this feat?

One reasonable answer, based on discoveries by Helen M. Blau of Stanford University, is that certain proteins (perhaps ones distinct from activators) in human cells may be capable of displacing nucleosomes from enhancers. It is also possible that genes a cell must turn on at one moment and off at another (like most genes in yeast) may contain enhancers that are kept perpetually at the ready—free of nucleosomes.

A good deal of work remains to be done before all the cellular activities of histones in yeast and other organisms are evident. Yet enough has been learned to overturn the old dogma that histone core particles are the molecular equivalent of unreactive spools. Our studies and those of others argue that the particles can both repress and activate genes. Individual histones, as well as specific domains within those molecules, account for the effects. It is tempting to speculate that certain regulatory proteins that collaborate with activators control genes by binding specifically to DNA and to key histone domains.

The challenge now is to learn more about the tasks nature has assigned to small segments within histone domains and to identify which regulatory proteins interact with those segments. Many investigators who would once have been uninspired by histones are now engaged in that effort. Exploration of chromatin structure has suddenly become an exciting new frontier in the study of gene regulation.

FURTHER READING

EXTREMELY CONSERVED HISTONE H4 N TERMINUS IS DISPENSABLE FOR GROWTH BUT ESSENTIAL FOR REPRESSING THE SILENT MATING LOCI IN YEAST. Paul S. Kayne, Ung-Jin Kim, Min Han, Janet R. Mullen, Fuminori Yoshizaki and Michael Grunstein in *Cell*, Vol. 55, pages 27–39; October 7, 1988.

NUCLEOSOME LOSS ACTIVATES YEAST DOWNSTREAM PROMOTERS IN VIVO. Min Han and Michael Grunstein in *Cell*, Vol. 55, pages 1137–1145; December 23, 1988.

YEAST HISTONE H4 N-TERMINAL SEQUENCE IS REQUIRED FOR PROMOTER ACTIVATION IN VIVO. Linda K. Durrin, Randall K. Mann, Paul S. Kayne and Michael Grunstein in *Cell*, Vol. 65, pages 1023–1031; June 14, 1991.

CHROMATIN AS AN ESSENTIAL PART OF THE TRANSCRIPTIONAL MECHANISM. Gary Felsenfeld in *Nature*, Vol. 355, No. 6357, pages 219–223; January 16, 1992.

SCIENTIFIC AMERICAN
In Other Languages

LE SCIENZE
L. 5,500/copy L. 55,000/year L. 75,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Piazza della Repubblica, 8
20121 Milano, Italy
Advertising correspondence:
A. Manzoni & C. S.p.A., Via Nervesa 21
20139 Milano, Italy

サイエンス
¥ 980/copy ¥ 10,753/year ¥ 13,393/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc., No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
600 Ptas/copy 6600 Ptas/year 7300 Ptas/(abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica, S.A.,
Viladomat, 291 6º 1ª, 08029 Barcelona, Spain

POUR LA SCIENCE
32 FF/copy 310 FF/year 395 FF/year/(abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou, 75006 Paris, France

Spektrum DER WISSENSCHAFT
10.50 DM/copy 111.60 DM/year 121.80 DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft
Verlagsgesellschaft mbH
Moenchhofstrasse 15, D-6900 Heidelberg
Federal Republic of Germany
Advertising correspondence:
Gesellschaft für Wirtschaftspublizistik GWP mbH
Kasernenstrasse 67, D-4000 Duesseldorf,
Federal Republic of Germany

科学
3.80 RMB/copy 45.60 RMB/year $48/(abroad)
Editorial, subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

В МИРЕ НАУКИ
2 R/copy 24 R/year $78.50/(abroad)
Editorial correspondence: MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow, U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc., 12224 Parklawn Drive
Rockville, MD 20852, USA

TUDOMÁNY
98 Ft/copy 1,176 Ft/year 2,100 Ft/(abroad)
Editorial correspondence: TUDOMÁNY
H-1536 Budapest, Pf. 338, Hungary
Subscription correspondence: "KULTURA"
H-3891 Budapest, Pf. 149, Hungary

العلوم
1 KD/copy 10 KD/year $40/(abroad)
Editorial, subscription, advertising correspondence:
MAJALLAT AL-OLOOM
P.O. Box 20856 Safat, 13069, Kuwait

ŚWIAT NAUKI
Zl. 22,000/copy
Correspondence:
Science Press Sp. zoo
00 789 Warszawa, ul. Sloneczna 35, Poland

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue, New York, NY 10017
Telephone: (212) 754-0550 Telex: 236115

Korean

TECHNOLOGY FOCUS 1992

KOREAN TECHNOLOGY FOCUS

by Oles Gadacz

Kim Jin-Hyun, the Republic of Korea's Minister of Science and Technology

Kim Jin-Hyun, Korea's Minister of Science and Technology, is well aware of the need to accelerate the national R&D effort. Economic planners here envision Korea joining the ranks of the G7 countries sometime in the early 21st century, but the nation currently faces slowed economic growth, and the commonly prescribed solution involves massive increases in R&D spending to upgrade the national technology level. "To become a self-reliant and advanced country," says Kim, "Korea must independently strengthen its industrial and technical capability." The fact that many multinational companies spend more on research and development than the Republic of Korea does lends urgency to Kim's crusade to increase R&D spending.

High wages and the lack of independent technological capability are cited as the chief reasons for Korea's eroding international competitiveness. During the 1980s, a decade of extraordinary growth for Korea, R&D spending rose from 0.58 percent to 1.91 percent of the gross national product. The government is now promising to raise the figure to 5 percent by 2001, which would match or surpass the level of the G7 countries. "The target," says Kim, "is the biggest in the world."

In 1991, against this background, the Korean government introduced the Highly Advanced National Project (HANP), aimed at selecting and developing strategic industrial technology. The strategy is clear: Technology, rather than science, should be emphasized, and limited national resources should be concentrated in critical areas. Among the specific goals of the project:

Semiconductors: the production of a 256M DRAM device by 1996, and the development of a 1G DRAM device by 2000.

ECONOMIC PLANNERS HERE SEE KOREA JOINING THE RANKS OF THE G7 COUNTRIES SOMETIME IN THE EARLY 21ST CENTURY.

Telecommunications: the development of asynchronous transfer mode by 1996, and the realization of an integrated-services digital network (ISDN) by 2000.

Consumer Electronics: the development of a high-definition television monitor by 1993, HDTV transmission and broadcasting ability by 1994, and a large-screen flat-panel display by 1997.

Fine Chemicals: the development of new antibiotic and germicidal agents by 1997.

Manufacturing: the development of computer-integrated manufacturing by 1996, and of intelligent manufacturing systems by 2000.

HANP is also reaching out to cultivate fundamental technology in the areas of machinery, biotechnology, the environment and energy, and it will support research into new materials for information services, energy and electronics.

The two organizations responsible for Korea's R&D programs are the Ministry of Trade and Industry (MTI) and the Ministry of Science and Technology (MOST). MOST is in charge of fundamental technology, and MTI oversees the development of industrial technology. Other governmental bodies, such as the Ministry of Energy and Resources and the Ministry of Communications, also participate in certain projects. Explicit policies and careful coordination guard against any overlapping of R&D efforts, and help to insure that the cost of large projects is shared.

HANP is supported by all government ministries, and will be carried out cooperatively by the nation's research centers, including universities, private companies and government-supported research institutes, such as Korea's two showcase R&D establishments, the Korea Institute of Science and Technology and the Korea Advanced Institute of Science and Technology.

Recognizing that lack of scientific manpower is another serious obstacle to economic growth, the government is stressing internationalization in the execution of HANP. MOST will invest 5 percent to 20 percent of the HANP research and development budget on international activities, and active international exchange programs for researchers will also be promoted. In a departure from past practices, foreign expertise will be solicited at the planning and evaluation stage of each project. In particular, active cooperation with advanced countries, such as the United States and Japan, is expected to begin in the near future.

Korea's success in achieving its goals, says Kim Jin-Hyun, is of importance to the United States. "We're a solitary country squeezed by giants—China and Japan," he says. "We desperately need U.S. help in science and trade, but unfortunately the United States is focused on the short-term balance sheet." Kim proposes a strategic alliance in which the United States contributes technology and Korea its manufacturing abilities.

"Korea is a strategically important partner to advancing U.S. interests in the region," Kim says. "I hope the United States considers geopolitical elements in considering its partnership with Korea."

In today's international environment, with borders becoming less relevant and global competition intensifying, Korea at last recognizes that she must focus her own resources on priority areas and take advantage of foreign resources as well. There is a growing realization that Korea's response to the challenges of the '90s will bear significantly on its destiny in the 21st century.

Oles Gadacz is a Seoul-based journalist specializing in the Korean automotive and electronics industries.

GOLDSTAR, SHINING BRIGHTLY

Goldstar Co., Ltd.,'s VISION2000 program sees the company emerging as one of the world's top five consumer-electronics companies by the end of the century. To reach the top, Goldstar has increased its expenditures on research and development from $247 million in 1989—equivalent to 6.5 percent of sales—to $700 million in 1992—7.5 percent. Management is committed to raising spending to 8.6 percent of sales, which would be about $1.4 billion, by 1995, and to 10 percent of sales by 2000. Goldstar's main Japanese rivals currently spend in the neighborhood of 8 percent of sales on R&D.

Goldstar's remarkable growth in the past was made possible by vigorous research and development for innovative products and technologies.

Korea's days of riding the coattails of Japanese technology are vanishing. "It's my wish to see us develop one innovative product during the '90s in the area of consumer electronics," says Dr. Kang In-Ku, executive vice president of Goldstar's Central Research Laboratory. "One product born of our own idea. I don't know what that's going to be, but I know we can do it. It needn't be expensive, but it will be a breakthrough in revealing a new attitude. It's going to take a lot of changes in management, and in the thinking of our engineers.

"We've already come very close to it. We came up with an idea, only to see the Japanese come out with the very same thing."

Trying to encourage innovative thinking in a corporate culture that has traditionally valued consensus is a major challenge. Kang cautions that change will be evolutionary, not revolutionary. "I don't expect to see an innovative product in the near future, but I see attempts being made to try something new," he says. "That willingness to try something new in itself represents a major change."

Goldstar has already made some organizational changes—results of a study done by the consulting firm of McKinsey & Co. a few years ago. "We know we have to create value for the customer, and we're moving toward market-oriented R&D," Kang says. "The most critical parts and components are being imported from Japan. We realize that we have to reduce this, and that this is hindering our R&D activities."

Kang faults Korea's educational system for another problem: a shortage of good researchers, though more than 150 of the 8,200 researchers currently employed by Goldstar have doctoral degrees. "The current system is discouraging innovation," Kang says.

Asked to list Goldstar's most significant technical achievements to date, Kang names high-definition television; the STAREX, a 6,000-line digital PBX that is now being exported; participation in the national central office-switch program; the independent capability to produce inverters and distributed control systems for industrial applications; and the development of navigational radar systems for use in leisure craft. Other areas of emphasis include erasable compact-disk players, home automation systems and liquid-crystal displays. The company is also concentrating on the development of flat-panel TV displays and digital and high-definition VCRs.

Goldstar is determined to be a key player in semiconductors. The company's plant in Chungju, which manufactures the 4M DRAM device, boasts the largest production capacity of any such plant in the world. A 16M DRAM device has been introduced, and efforts are currently under way to develop 64M DRAM devices; a prototype of the latter is expected to be ready in 1993.

The multibillion-dollar investment in DRAM research and production has helped Goldstar close the gap with the Japanese. The company is now only several months behind the industry leaders, and sales have been encouraging: According to Dataquest, Goldstar was 11th in the world in DRAM sales in 1991.

The company's R&D spending is currently being divided among the Central Research Lab, the Information Systems Lab, the Production Engineering Research Lab and the Quality Assurance Lab. Goldstar has five facilities in other countries, including Goldstar Technology, established in 1984 in Sunnyvale, California, and the Goldstar Japan Co., Ltd., in Tokyo. At the Goldstar North America Lab Inc., in Chicago, the focus is on HDTV. The Chicago lab was set up in 1991, when the company acquired a 5 percent stake in Zenith Electronics Corp. The two companies plan to cooperate closely in HDTV research. Zenith's HDTV system, developed in cooperation with AT&T, has a good chance of being selected by the Federal Communications Commission as the United States' national standard. Should that happen, Goldstar, instead of trying to catch up, will find itself on the inside track.

> A MULTIBILLION-DOLLAR INVESTMENT IN SEMICONDUCTORS HAS HELPED CLOSE THE GAP WITH THE JAPANESE.

TAEJON EXPO '93
FINALLY, THE WORLD PAYS A VISIT

Based on the concepts of light, science and the universe, the Grand Light Tower is a three-dimensional feature representing the past, the present and the future.

In 1893, Korea took its first hesitant steps into the world, sending an exhibit of simple handicrafts to the international exposition in Chicago. Next year, exactly a century later, it will be Korea's turn to play host.

Ten million visitors are expected during the three-month run of EXPO '93, which will open on August 7 in Taejon, Korea's sixth-largest city.

ONE OF THE GOALS OF EXPO '93 IS TO CREATE ENTHUSIASM FOR SCIENCE AMONG KOREA'S YOUNG PEOPLE.

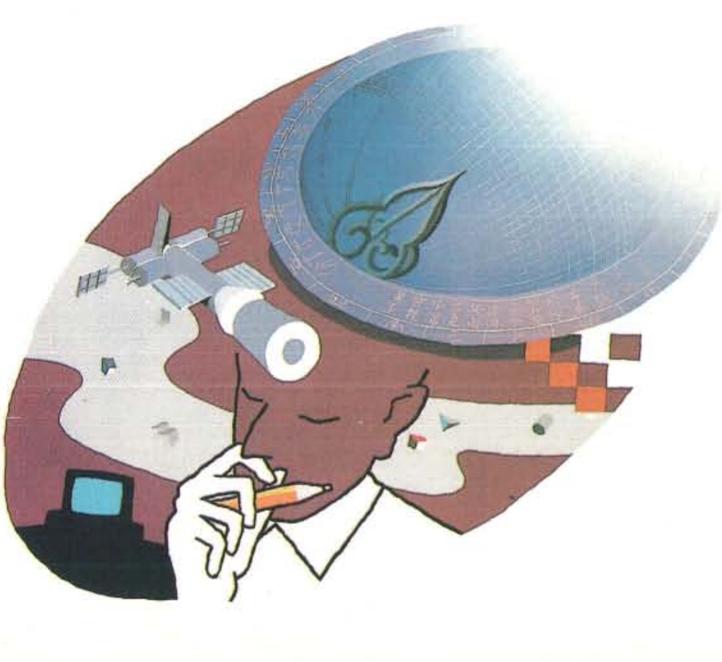
The theme will be "The Challenge of a New Road to Development"; under that banner, Korea will be using the event to spur worldwide efforts to find solutions to the problems of industrialization. Another goal is to create enthusiasm for science and technology among Korea's young people.

Visitors will learn about foreign countries and organizations in the One World Zone. The Cultural Innovation Zone will feature exhibits on Korea's cities and provinces, and pavilions covering culture and art, housing, and the environment. The Industrial Prosperity Zone will examine the Korean economy.

Corporate Korea will flex its technological muscles in the 15 theme pavilions. The Lucky-Goldstar Group, for example, will build the Electronics and Computer Pavilion. The Hyundai Group will erect the Transportation Pavilion, and Korea's number-two automaker, the Kia Group, will be responsible for the Automobile Pavilion. Pohang Iron & Steel Co. will build the Materials Pavilion. The Hanjin Group, operators of Korean Air, will run the Future Aerospace Pavilion. And Korea Telecom will examine the future of telecommunications at the Information and Communications Pavilion.

EXPO '93

One Site for the World, One Sight of the Future



Imagine the possibilities of the future

Think, if you will, about the limitlessness of the future. How technology has taken us from room-sized computers to ones no bigger than a notebook. From horse and buggy to space flight. From horsepower to solar power. All in just one century.

Between August 7 and November 7, 1993, visitors at Taejon Expo '93 will be able to do just that, exploring the frontiers of technology, environmental preservation, and resource conservation as well as rich cultural heritages from around the world.

Start imagining the possibilities of the future. Attend Taejon Expo '93 and join the human race as it rushes toward the 21st century.

For more information and a full-color brochure on Taejon Expo '93, write or call the following numbers today.

The Taejon Int'l Exposition, Organizing Committee

159 Samsung-dong, Kangnam-ku, Seoul 135-090, Korea
Phone: (02) 551-4870/1 Fax: (02) 551-1316

IN EVERY GOLDSTAR PRODUCT THE MOST IMPORTANT COMPONENT IS YOU.

Before even the first line is drawn, our design team works with you in mind. The result is a portable powerhouse you'll feel more than comfortable with. As light and compact as the average briefcase, it gives you the flexibility to take it everywhere. And the high-resolution LCD will brighten any presentation. Just one more example of how Goldstar harnesses high technology to work for you. A practice that has placed us at the forefront of Korean electronics and is winning applause all over the world.

GoldStar
Technology with the human touch.

Goldstar Co., Ltd. Yoido P.O. Box 335, Seoul, Korea Tel: (02) 787-1114 Fax: (02) 787-3400 Tlx: GSRADIO K23751/3

SAMSUNG ELECTRONICS—A FOLLOWER NO LONGER

Samsung Electronics, which relied on licensed technology to spur its growth over the past two decades, is now blazing a path of its own. Samsung has built up formidable R&D power across a broad spectrum of home and industrial electronics products. But nowhere does the company's star shine more brightly than in the area of semiconductors, and particularly memory devices.

In the early 1980's, ignoring the skeptics, Samsung decided that it would target simple types of semiconductors, such as DRAM and SRAM devices. That decision has been paying off handsomely. It is in this intensely competitive field that Samsung has succeeded most quickly in narrowing the technology gap with the Japanese, the industry leaders. Samsung's state-of-the-art 16M DRAM device was introduced in August 1990. A prototype of the next-generation device, a 64M DRAM, is scheduled to be unveiled sometime in 1993. According to Dataquest, Samsung stood second only to Toshiba in international DRAM sales in 1991, and in fourth place in overall sales of memory devices.

Samsung is aiming even higher. Drawing on the resources of the Samsung Group, a conglomerate with a consolidated turnover of $43.9 million in 1991, the company captured 13th place in the world semiconductor industry in 1991 on sales of $1.6 billion. Samsung is determined to make the top-10 list of electronics companies by 1996, and the top five by early next century. Industry analysts think that its chances are good.

Samsung's R&D outlays are commensurate with its goals. The company's R&D spending leads the nation, having jumped from $526 million in 1990 to $626 million in 1991—an increase from 7.7 percent of sales to 8.4 percent. In this year's budget, $838 million, or 9.3 percent of sales, is earmarked for research and development, and Samsung expects to spend $1.1 billion on R&D in 1993.

Samsung Electronics leads the development of the Korean-made TDX-10.

Staffing at research centers is being expanded just as aggressively. In 1991, Samsung's R&D centers employed 8,880 researchers, accounting for 20 percent of all the company's employees. Samsung expects the number to grow to 12,150 by 1993.

Attracting high-caliber researchers is easier than it used to be. The brain drain of the '70s and '80s, which saw many of Korea's best scientific minds leave the country because of the backward state of science, has reversed in recent years. Many of these scientists have returned home, drawn by the chance to help Korea make the critical transition to technology-intensive manufacturing. Samsung now employs 195 holders of doctoral degrees, many of whom studied at prestigious American and European universities. Samsung estimates that the number of researchers in Korea with Ph.D.s will grow at the astonishing rate of 38 percent over the next several years.

> SAMSUNG'S R&D SPENDING LEADS THE NATION, AND RESEARCH STAFFS ARE BEING EXPANDED JUST AS AGGRESSIVELY.

More than 3,000 of Samsung's researchers—more than a third of the total—work in the electronic home-appliance division, the company's major generator of revenue. Recent milestones include the development of an 800-line-resolution television, a video tape recorder with better than 400-line resolution, a digital audio tape player, a laser-disk player, a washing machine with fuzzy logic, and a five-inch liquid-crystal-display television.

The division has more than 300 projects in progress, including high-definition television, a mid-size LCD television, a digital VCR, a second-generation security robot and an artificial-intelligence robot. Among the longer-term projects are HDTV with three-dimensional capability and ISDN compatibility, an ISDN-compatible VCR, a large-screen LCD television and mobile robots.

One of the largest and most important projects in Samsung's information telecom division is its work on the TDX-10, a digital central-office switch with 100,000-line capacity, which is being developed and manufactured jointly by four Korean companies for Korea Telecom (KT), the state-owned public common carrier. Samsung has poured more than $138 million into the project, and has earned first place in KT's evaluation of the four companies. Employing more than 1,000 32-bit microprocessors for its distributed-architecture design, the TDX-10 places Korea in the elite club of nine nations that manufacture high-capacity digital central-office switches.

Samsung's small but growing line of TDX central-office switches is finding customers in the Philippines, Latin America, Poland and the Commonwealth of Independent States. Most promising of all, and potentially worth hundreds of millions of dollars, the CIS and China recently entered coproduction agreements with Samsung to modernize their telecommunications infrastructures with TDX systems.

Another of Samsung's recent accomplishments is the independent development of Korea's first G4-class fax machine. The company is moving steadily toward the realization of the integrated-services digital network through its research into a digital wireless PABX, broad-band ISDN switching systems, satellite fax and an automatic-interpretation telephone.

Singing Caterpillars, Ants and Symbiosis

With siren songs and chemical lures, some butterfly caterpillars attract ants to serve as their guardians. Those same ants are usually involved in symbioses with other insects and plants as well

by Philip J. DeVries

As anyone who has picnicked atop an anthill can attest, ants vigorously defend the food they find and strongly discourage poaching on their territory by other species (human or otherwise). Yet some insects, including a breathtaking variety of butterfly caterpillars, can not only trespass safely but enter into mutually beneficial partnerships with ants. Such relationships are intriguing examples of symbiosis, in which two or more species may live and interact intimately. Because symbioses offer great clues for understanding the complex interactions between multiple species, they are of special interest to evolutionary ecologists, who are concerned with how and why organisms have evolved their particular traits and behaviors.

The ability to form symbioses with ants and to exploit their pugnacious characteristics is well known in two major groups of organisms: plants and herbivorous insects belonging to the orders Homoptera (aphids, cicadas and related insects) and Lepidoptera (butterflies, skippers and moths). For all these organisms, the presentation of food in the form of tasty secretions seems to be fundamental to maintaining the association with the ants. Plants may provide secretions through extrafloral nectaries on their leaves. Ants drawn to these nectaries defend the plants against herbivorous insects. Similarly, insects can offer secretions to ants through specialized organs. Ants attracted to the honeydew secretions produced by aphids, for example, seem to protect those insects against predators.

Myrmecophily, the ability to associate symbiotically with ants, has evolved among only two families of butterflies: the Lycaenidae (which are known as hairstreaks and blues and are found everywhere in the world) and the Riodinidae (or metalmarks, which are found almost exclusively in the American tropics). Together these families make up what are commonly called the lycaenoid butterflies. Lycaenoids are seldom noticed—they have wingspans of less than two inches—but they constitute 40 percent of the more than 13,500 known butterfly species and encompass a staggering diversity of colors and patterns. As delightful to the eye as lycaenoid butterflies may be, it is their caterpillars that dazzle the imagination of evolutionary biologists because many of them have specialized organs that mediate their symbioses with ants.

During the past 20 years, this relationship has been illuminated by Christopher B. Cottrell of the Tobacco Research Board in Zimbabwe, Konrad Fiedler of Maximilians University in Würzburg, Ulrich Maschwitz of Goethe University in Frankfurt am Main, Naomi E. Pierce of Harvard University and Jeremy A. Thomas of the Institute of Terrestrial Ecology in England, who have studied members of the Lycaenidae family. These researchers have shown that the symbioses between lycaenid caterpillars and ants can range from the mutualistic, in which both species benefit, to the parasitic, in which one benefits at the expense of the other. In some instances, the symbioses lead to complicated life cycles for one or both species.

Until recently, the only detailed work on symbioses between riodinid caterpillars and ants was that done more than 20 years ago by Gary N. Ross of Louisiana State University at Baton Rouge. For a very long time, therefore, our grasp of the evolution and ecology of butterfly-ant symbioses was based almost entirely on studies of the family Lycaenidae. Recent research on riodinid caterpillars by myself and others, however, yielded comparative information that has helped reinterpret the evolution and consequences of symbiotic associations with ants. With that knowledge has come a fresh appreciation of the pivotal ecological role that some ant species play.

My interest in caterpillar-ant interactions started some years ago in Brunei, Borneo, where I first observed interactions between ants and a lycaenoid butterfly. At the time, most of my work centered on other aspects of butterfly biology, but from this incidental observation grew a fascination with myrmecophilous (ant-loving) caterpillars. During the past seven years, much of my research has focused on exploring the symbioses between ants and caterpillars of the riodinid butterflies in Central and South America.

The riodinid butterfly that I have studied in greatest detail is *Thisbe ire-*

PHILIP J. DEVRIES has broad interests in evolution, ecology, comparative biology and natural history, but his special expertise is in plant-animal interactions, particularly those involving insects. He is a senior research fellow in zoology at the University of Texas at Austin and a scientific associate of the Center for Conservation Biology at Stanford University. After graduating from the University of Michigan in 1975, DeVries began graduate work in zoology at the University of Texas. He received his doctorate in 1987. His professional honors include a MacArthur Foundation fellowship.

SYMBIOTIC PARTNERSHIP between this caterpillar and ants is beneficial to both species. A variety of adaptations, including attractive songs or acoustic signals, enable the caterpillar to exploit the territoriality and the food-gathering instincts of the ants. This caterpillar is sipping sweet fluid from an extrafloral nectary on a plant while ants tend it.

nea, which lives in a variety of tropical forest habitats between Mexico and Brazil. It is typical of riodinid caterpillars that form symbioses with ants. A female *Thisbe* deposits individual eggs on sapling trees of the genus *Croton;* after caterpillars emerge from these eggs, they feed on the tree. Ants are drawn to *Croton* because at the base of each leaf is an extrafloral nectary. (Such nectaries are common features of many tropical plants.) The ants patrolling the trees are the same ones with which the caterpillars form symbioses.

When I began my study of *Thisbe* in Barro Colorado, Panama, all that was known about the caterpillars was that they fed on *Croton* and were typically found in association with ants. The logical first step was to see what would happen to the caterpillars without ants. I therefore removed all the insects from several populations of *Croton* saplings and then smeared the base of each plant with a sticky resin. This arrangement would deny all crawling insects, such as ants, access to the trees but would not prevent female butterflies from laying eggs there. On half the trees, however, I placed a stick bridge over the resin to allow ants to return.

For the next 10 months, I conducted a weekly census of ants and caterpillars on all the trees. The results showed that the trees with ants on them accrued more caterpillars than did the trees from which ants were excluded. A likely explanation for this pattern seemed to be that on the plants lacking ants, flying predators were removing the caterpillars.

Of the natural predators that caterpillars face, the social wasps are especially significant, particularly in the tropics. Social wasps spend a large proportion of their adult lives searching vegetation for caterpillars. When a female wasp finds a caterpillar, she kills it with her sting, cuts up the body and carries the meat back to her nest to feed the hungry wasp grubs.

To test whether ants protected caterpillars from wasps, I placed two potted plants in an area where wasp nests were abundant and allowed ants to crawl on only one of the plants. I then placed a caterpillar on each plant and timed how long it survived. As eventually became obvious, in the absence of ants, caterpillars did not remain on plants for very long: often within minutes, wasps killed the caterpillars and carried them away. In contrast, if ants were present, they vigorously defended the caterpillars against wasp attacks. Wasp predation therefore explained why I found fewer caterpillars on the plants from which ants had been excluded.

These simple experiments showed that ants performed a valuable service to *Thisbe* caterpillars by protecting them from predators. That observation raised a further question about the symbiosis: what did the ants gain in return for their efforts? Or to put it another way, how do the caterpillars induce ants to engage in a risky defense of an organism that is not part of their colony?

The answer rests, in part, with a suite of specialized organs found on the caterpillar. Ants generally ignore young *Thisbe* caterpillars that are still in their first and second instars, or developmental stages. Only on molting and entering the third instar do the caterpillars undergo a dramatic morphological change that makes them highly attractive; thereafter, ants attend them constantly until the caterpillars begin their metamorphosis into butterflies.

Caterpillars in their third and later instars have three sets of so-called ant organs that are important for maintaining the tending activities by the ants. The most noticeable are a pair of extrudable glands called nectary organs. Located on the caterpillar's posterior segments, the nectary organs look like the fingers of a rubber surgical glove. When an ant strokes the posterior area of a caterpillar with its antennae, these organs extrude from the body and secrete a drop of clear fluid at their tips that the ant drinks eagerly. The organs

THISBE IRENEA **is one of many butterflies that as a caterpillar enters into a symbiosis with ants. Myrmecophily, "the love of ants," has evolved in butterflies only within the families Riodinidae and Lycaenidae.**

then withdraw back into the body. The ants, however, are so captivated by the secretion that they stroke the caterpillar relentlessly to solicit it. I estimate that the ants attending a *Thisbe* caterpillar seek more of the secretion at least once every minute.

These same ants also obtain secretions from the extrafloral nectaries of *Croton* trees, but they seem to prefer tending the caterpillars to tending the nearby plant nectaries. The late Irene Baker of the University of California at Berkeley and I found that the caterpillar secretions are quite different from those offered by the *Croton* plants. In effect, the secretion of the caterpillars is haute cuisine. True, the extrafloral nectar is a 33 percent mixture of various sugars, whereas the caterpillar secretion has almost no detectable sugar, but the secretion contains much higher concentrations of amino acids. Ants can obtain a meal from the caterpillars that is more nutritious than what they can get from the plant nectaries, even if it is not so sweet.

The caterpillars, on the other hand, are partial to the extrafloral nectar. When *Thisbe* caterpillars are not feeding on leaves or walking about on the *Croton* plant, they typically rest with their heads over the extrafloral nectaries. A series of experiments in which caterpillars were grown with and without access to extrafloral nectar showed that caterpillars did indeed drink the nectar and that those drinking it grew faster than those that did not. Extrafloral nectar, along with the other leaf tissue that serves as food, therefore contributes substantially to the developmental rate of *Thisbe* caterpillars.

Drinking extrafloral nectar appears to be widespread among the riodinid caterpillars that form symbioses with ants. Myrmecophilous caterpillars often feed on plants that have extrafloral nectaries. Conversely, caterpillar species that do not enter into symbioses with ants feed more frequently on plants lacking nectaries. By usurping the ant guards of the plant for their own protection and feeding on the young leaf tissues, the myrmecophilous riodinid caterpillars exploit the plant-ant symbiosis—adding insult to herbivory, so to speak.

Studies of the nectary organs solved only part of the caterpillar-ant symbiosis puzzle. Ants are selfless automatons that provide food, defense and brood care for their colony as a whole. Yet individual ants sometimes stay with *Thisbe* caterpillars for a week or more. Instead of spending time protecting members of a different species, why don't the attending ants immediately return to their nest with the caterpillar secretion, as they would with plant nectar or other delicacies?

Other ant organs on the caterpillars bear on that question. The caterpillars of *Thisbe* and other riodinid species have a pair of tentacle organs, brush-tipped glands just behind the head, that appear to exert a chemical influence on ant behavior. When these organs emerge from the body, the attending ants near the caterpillar almost immediately snap into a defensive posture, their mandibles agape, abdomen curled under the body. I found that when the ants assumed this posture, moving a small piece of wood or thread near the caterpillar precipitated an aggressive attack: the ants would rush at the object, bite it and attempt to sting it.

These observations of *Thisbe* and other riodinid species suggest that the tentacle organs discharge some chemical similar to an ant alarm pheromone—a substance that the ants would use among themselves to signal an attack on the colony. Unfortunately, the chemical nature of the emissions from the tentacle organs is still unknown, so it is impossible to say how they compare with actual ant alarm pheromones. It does seem clear, however, that the function of these organs is to seize the attention of ants and to help keep it focused on the caterpillars.

A third type of ant organ was discovered in 1926 by entomologist Carlos T. Bruch of the National Museum of Natural History in Buenos Aires, who provided the first description of a myrmecophilous riodinid species found in Argentina. In addition to the extrudable nectary and tentacle organs, Bruch found a pair of tiny, movable, rodlike appendages that sprang from the front of the first thoracic segment and projected over the head. Such appendages were then unknown on any other caterpillar.

Forty years later in Mexico, Ross described similar appendages on a different riodinid species. Ross named these appendages vibratory papillae and suggested that their motion might convey vibrations to ants. It is now known that most myrmecophilous riodinid caterpillars have vibratory papillae as ant organs.

My observations indicated that the vibratory papillae on *Thisbe* caterpillars functioned as Bruch and Ross had described. I also noticed, however, that while the vibratory papillae were in motion, the caterpillars moved their heads in and out. I was struck by the similarity of these head movements to those of long-horned beetles, which produce an audible sound when they move their heads in and out. (Predators seem to drop the beetles when they hear the squeaks.) Even though I could hear nothing coming from *Thisbe* caterpillars, I became convinced that they were producing sounds.

A glimmer of the truth began to ap-

pear when I examined the structures with a scanning electron microscope at the University of Texas at Austin. Under magnification, the vibratory papillae clearly showed sharp concentric rings along their shafts. The top of the head, which the papillae touched when they vibrated, bore small bumps or granulations that looked like guitar picks. Side by side in those micrographs, the vibratory papillae and the head were strongly reminiscent of a guiro, a Latin American percussion instrument that is played by sliding a wooden stick across the grooves in a carved gourd. Thus, both the morphology and the behavior of the caterpillar suggested that these papillae and the head acted jointly as a sound-producing organ.

No one had ever heard a caterpillar make a sound, however, so this crazy idea needed to be tested. I consequently returned to Panama with an exceptionally sensitive microphone and amplifier. On the very day of my arrival, I discovered that the instruments did indeed allow me to hear and record calls from the caterpillars. Their low-amplitude calls were audible only when the microphone was touching the body of a caterpillar or the surface on which it stood. The calls therefore travel through the solid substrate rather than through the air, which partially explained why no one had ever heard caterpillar calls during normal observation.

By making recordings of individual caterpillars and then removing their vibratory papillae, I learned that the caterpillars could produce calls as long as they had at least one papilla. As analyses of the recordings revealed, the calls were strongest when two vibratory papillae were present and weaker by almost half when one was removed. Caterpillars without any vibratory papillae were completely mute even though they continued to move their heads. Because these papillae are replaced when a caterpillar sheds its skin, the mute caterpillars regained their "voices" when they molted to the next instar. (That fact was experimentally useful because individual caterpillars could be used several times, and they acted as built-in controls.) Taken together, these observations verified that the vibratory papillae, the head granulations and the head motions worked literally in concert as components of a call-producing system.

The ability to make caterpillars mute permitted me to explore the role of caterpillar calls in the formation of associations with ants. I placed mute and call-producing caterpillars on the same plant and counted how many ants tended each of them over time. The results indicated that the calls do keep ants in the vicinity of the caterpillars: the calling caterpillars accumulated significantly more ants. It is easy to imagine that the calling caterpillars would also have been better protected against predators than their silent counterparts because they could attract more ants.

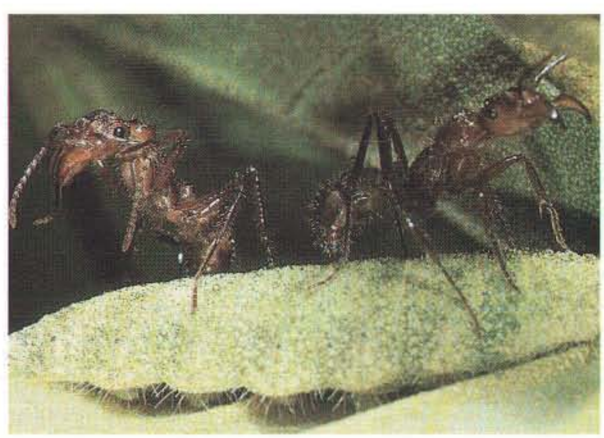
Just as the chemical stimulus from the tentacle organs seems to mimic a pheromonal signal among ants, the calls of the caterpillars seem to parallel the auditory communications of ants, at least in some respects. When ants find a food resource or when they are alarmed, they create vibrations that travel through the substrate on which they stand and attract their nest mates. Many varieties of ant produce vibrations by tapping their abdomen on the substrate, but others have well-developed sound-making organs. The calls of ants and caterpillars have approximately the same frequencies and pulse rates.

These similarities suggest that the ability of caterpillars to call and the characteristics of the calls themselves evolved through natural selection by ants. In other words, the ants determined which caterpillars survived by ignoring strange calls and responding to antlike calls.

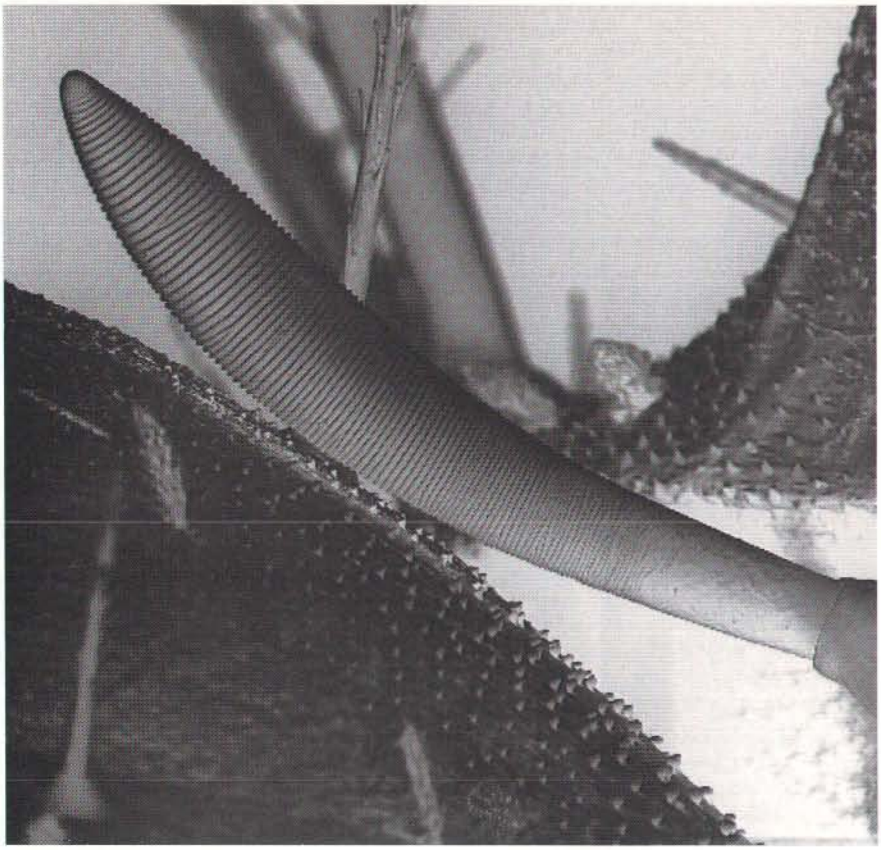
That behavior is noteworthy because it violates what biologists have generally understood about insect communication systems—that calls evolve in response to direct selection on sexual or defensive traits. For example, the chirping of a male cricket summons potential mates and warns other males away from its territory. Vigorous males proliferate faster than do their sluggish brothers because their calls attract larger numbers of receptive females. The calling ability of crickets and the characteristics of their calls thus evolve directly in response to selection by females of the same species.

In caterpillar and ant symbioses, however, the calls produced by one species have evolved in response to selection by an unrelated species. The evolution of the caterpillars has been guided by selection acting on completely different characteristics in ants. The caterpillar-ant system therefore provides a new arena for the study of insect communication, particularly in other insect species that form symbioses with ants.

Caterpillar-ant symbioses are typically facultative rather than obligate—that is, the two species benefit each other but neither is absolutely dependent on the other for survival. Indeed, the involvement of species in a symbiosis can be quite flexible: depending on the habitat and locality, a particular species of caterpillar may associate with a variety of ant species. For



SYMBIOTIC ANT BEHAVIORS are mediated by specialized ant organs on myrmecophilous caterpillars. At their posterior end, the caterpillars have extrudable nectary glands, which secrete an appetizing fluid for the ants to drink (*left*). At their anterior end, the caterpillars have tentacle organs, which release a chemical similar to an ant alarm signal. The ants immediately assume a defensive posture (*right*) and will protect the caterpillar from predators, such as wasps.



VIBRATORY PAPILLAE extending over the caterpillar's head from the first segment of the thorax help the insect to produce an acoustic call that attracts ants. Each papilla is a rod covered by concentric grooves. When the caterpillar moves its head in and out, microscopic granulations or bumps on the head slide across the grooves, producing vibrations (*left*). In form and function, the vibratory papillae and the head granulations resemble a guiro, a musical instrument that is played by sliding a rod across a carved gourd (*above*).

example, within its geographic range, *Thisbe* caterpillars may be found with one species of ant in Costa Rica and another species in Belize. Even within a single forest, different ant species often tend the caterpillars in different habitats.

These variations suggested to me that the calls of the caterpillars do not attract particular species of ants. Instead they seemed to be general attractive signals to which many species could potentially respond. To explore that idea, I needed to study the calls of caterpillars involved in obligate, species-specific symbioses.

The best-documented cases of obligate caterpillar-ant symbioses involve the European lycaenid genus *Maculinea,* which has caterpillars that are parasites of specific *Myrmica* ants. The work of Thomas and his co-workers in Europe has unraveled an amazingly complex set of *Maculinea* life histories, only part of which I will describe. When these caterpillars reach their third instar, they fall off their food plant and are picked up by a *Myrmica* ant. The caterpillar is then carried into the ant's nest, where it turns carnivorous and eats ant larvae. Although any *Myrmica* ant species will carry any species of *Maculinea* caterpillar into the nest, each species of ant allows only one caterpillar species to complete its life cycle within the nest; the ants eventually kill any other type of predatory caterpillar. As one might imagine, the selection pressure on *Maculinea* caterpillars to form symbioses with the appropriate species of *Myrmica* ant is extremely strong.

I decided to investigate whether the calls of these caterpillars were more similar to those of their specific hosts than to those of other ants. In collaboration with Thomas and Reginald B. Cocroft of Cornell University, I recorded and analyzed the calls produced by various *Maculinea* caterpillars and *Myrmica* ants. Each species of caterpillar and ant produced distinctive calls, but we found no obvious evidence that the calls of the caterpillars had evolved to resemble those of their obligate ant hosts.

Our observations supported the idea that caterpillar calls attract ants in general, not one species in particular. Once the ants are in attendance, the calls act along with other ant organs to maintain a constant guard of ants. In the *Maculinea-Myrmica* systems, chemical cues are probably most important for governing the species-specific symbioses.

Because the symbioses between caterpillars and ants are so remarkable, I became curious about how these structures and the phenomenon of myrmecophily evolved in caterpillars. Among the thousands of butterflies and moths in the order Lepidoptera, the ability of caterpillars to form symbioses with ants is unknown outside the lycaenoids. The Riodinidae and Lycaenidae families show many parallels in their biology. Myrmecophilous caterpillars in both groups produce food secretions, chemical signals and acoustic calls. Conversely, nonmyrmecophilous species lack all or some of these behaviors and the associated ant organs. They are always mute, for example, even if they are closely related to calling, myrmecophilous species.

Taken together, these patterns appear to refute an old hypothesis about the evolution of caterpillar-ant symbioses and to support a new one. Forty years ago, in a review paper on myrmecophilous Lepidoptera, Howard E. Hinton of the University of Bristol, England, made three explicit points about the evolutionary history of these species. First, he maintained, myrmecophily evolved just once in butterflies; second, myrmecophily is a primitive trait held over from that one event; third, the ability to form symbioses with ants has been lost in some lineages of riodinids and lycaenids.

A comparison of the riodinid and lycaenid caterpillars with each other and of both types with other Lepidoptera suggests an evolutionary history that differs greatly from what Hinton envisioned. Although riodinid and lycaenid caterpillars have ant organs that secrete food and chemical signals, these organs develop on different body seg-

ments in each group. Lycaenid caterpillars produce calls, but they do not have vibratory papillae; the ant organs with which they call are still unknown. The ant organs of the riodinids and the lycaenids are therefore analogous but not homologous—they are similar but not evolutionarily related to each other. Comparative morphology alone thus suggests that myrmecophily evolved at least twice in butterflies: once in the lycaenids and once in the riodinids.

Moreover, the specific mechanisms mediating the symbioses may have evolved at least three times. Caterpillars of the riodinid genus *Eurybia* make calls and form symbioses with ants, but like lycaenids, they do not possess vibratory papillae. That fact implies that the ability of butterflies to produce calls must have evolved at least twice among the riodinid caterpillars alone.

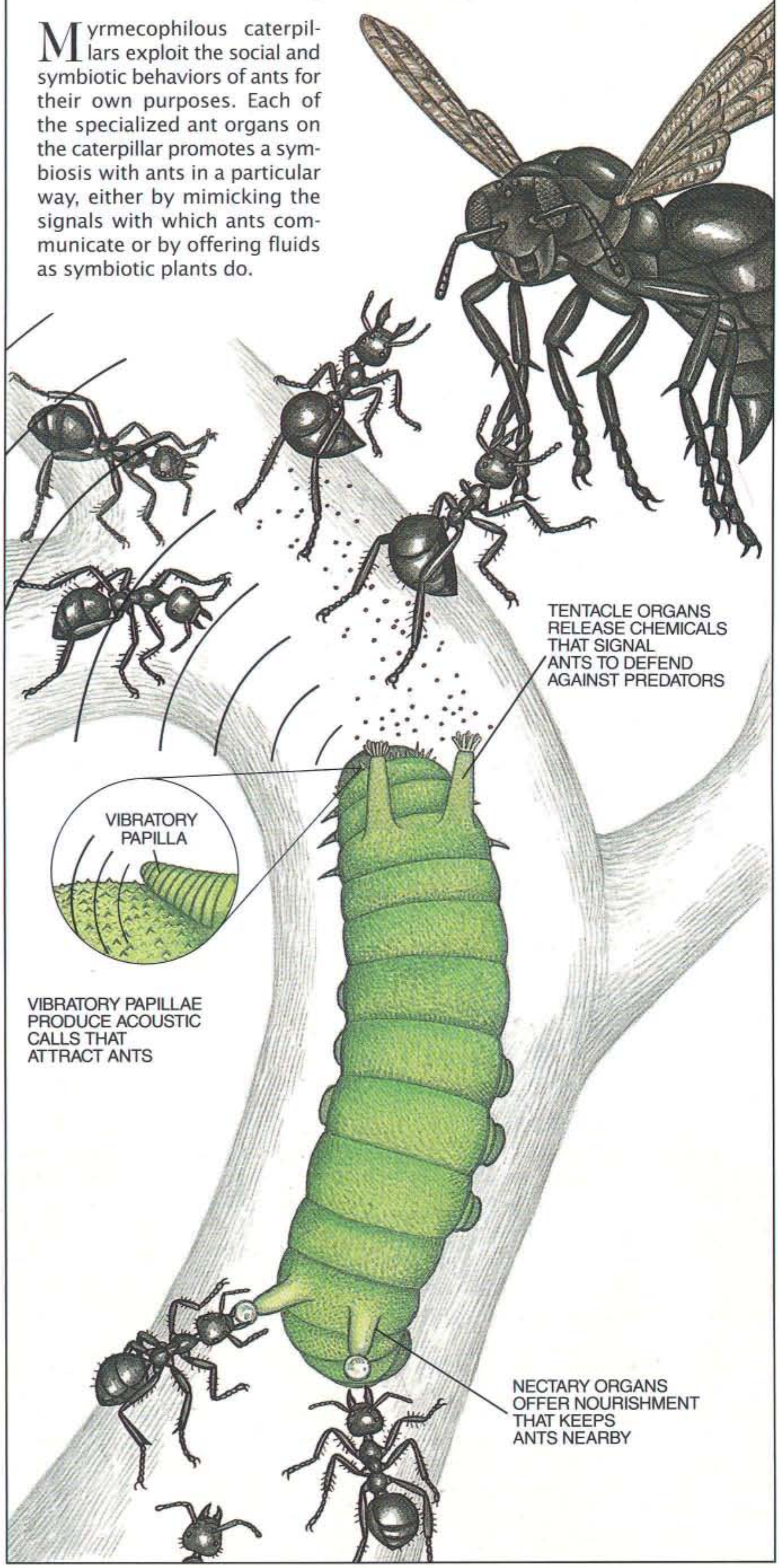
The nonmyrmecophilous species are also instructive about the evolution of caterpillar-ant symbioses. The vast majority of butterfly and moth caterpillars bristle with spines, hairs or special behaviors that keep ants away. Lycaenoid caterpillars that do not associate with ants also have long hairs and lack ant organs. Overall, the species of riodinids and lycaenids that do not establish symbioses with ants are more similar to most Lepidoptera than to their own myrmecophilous cousins. This fact, too, suggests that myrmecophily evolved in lycaenoid butterflies at least twice and that the trait appeared relatively recently.

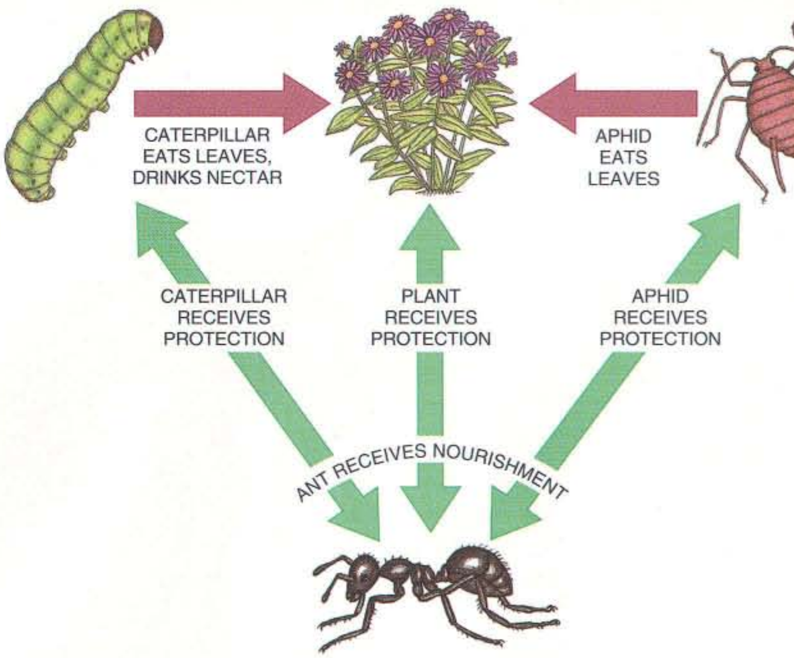
A look at the ants helps to explain some aspects of the development of the symbioses. Despite their apparent outward similarity, ant species are finely tuned specialists. For example, some ants have expandable abdomens for storing fluids and spend their lives hanging inside the nest as living canteens; army ants specialize in hunting and capturing certain kinds of arthropods; leaf-cutting ants are expert at chopping leaves into mulch for their underground fungus gardens. The tendency toward specialization among ants is so well established that species are often divided into four general categories by their diets: predators, scavengers, seed eaters and other herbivores, and harvesters of secretions.

Most studies on the evolution of myrmecophily have considered all ant species as exclusively either predators of caterpillars or potential mutualists with them. To the contrary, however, my observations showed that relatively few of the more than 100 ant species at my study site in Panama either cared for or preyed on caterpillars. When I of-

How a Caterpillar Manipulates Ants

Myrmecophilous caterpillars exploit the social and symbiotic behaviors of ants for their own purposes. Each of the specialized ant organs on the caterpillar promotes a symbiosis with ants in a particular way, either by mimicking the signals with which ants communicate or by offering fluids as symbiotic plants do.





INTERLOCKING SYMBIOSES are common because few species of ants tend other organisms. In any one habitat, plants, caterpillars and other insects, such as aphids, are all likely to engage in symbioses with the same ant species. Ironically, the sharing of symbiotic ants may allow these myrmecophilous insects to feed on plants that the ants otherwise protect from herbivorous insects.

fered riodinid and lycaenid caterpillars to an assortment of ants, a few species killed the caterpillars and a few tended them, but most simply ignored them.

What unified all the ant species that tended caterpillars was their feeding ecology: much of their time was spent actively harvesting the secretions produced by insects in the order Homoptera and by plant extrafloral nectaries, in addition to those of butterfly caterpillars. The pattern found locally in Panama soon proved to be pertinent on a much larger scale. Less than 10 percent of all the ant genera in the world associates with butterfly caterpillars, and this minority of ants also tends other secretion-producing insects and plants with extrafloral nectaries. Ant species that are general predators, specialist predators of arthropods, or herbivores and seed eaters are conspicuously absent from the list.

In short, regardless of the geographic locality—whether it be a rain forest in Ecuador, the Serengeti plains, a chalk down in England or Central Park in New York City—the ants that tend caterpillars also protect other secretion-producing insects and plants at the same site. In any one habitat, caterpillars, homopteran insects and plants may all establish symbioses with ants, but if they do, they will share the same ant symbionts.

The observed patterns suggest that historically only a small fraction of the total ant diversity may ever have been involved in the evolution of symbioses with caterpillars. In all probability, these few ants favorably influenced the evolution of insects and plants that had the ability to produce secretions. Once secretions became important in the diet of particular ants, any other insect or plant capable of producing secretions could have easily garnered ants as bodyguards, thereby promoting the evolution of new symbioses with ants.

The dynamics of the symbioses between butterfly caterpillars, other secretion-producing insects, plants and certain ants bring to mind a further evolutionary concept. Because myrmecophilous insect herbivores and plants use the same ants as mobile defenses, they may be competing for the ants' attention. For example, when *Thisbe* caterpillars draft the ants on *Croton* plants as guards, they not only succeed in gaining their own protection, they also circumvent the defense of the plant: *Thisbe* caterpillars can feed on *Croton*, whereas most other herbivorous insects cannot. The caterpillars have inserted themselves into the ant-plant symbiosis and subtly undermined it. It is possible that the evolution of myrmecophily has permitted herbivorous insects to invade and exploit symbioses between plants and ants.

The study of interacting species provides new perspectives on the natural world and forces us to look at symbiotic associations in a much more dynamic context. My work began as a simple description of the interaction between *Thisbe* caterpillars and ants and led to research on the specialized ant organs that mediate their symbioses. As typically happens in ecological studies, a knowledge of other plant and insect species soon became essential for even a rudimentary understanding of the *Thisbe*-ant system. The interactions and parallels between symbioses resulted in findings relevant to understanding insect-plant dynamics, predator-prey interactions, insect communication systems and the evolution of symbioses involving insects, plants and ants.

The study of ants and myrmecophilous species has shown me a few of the many interactions that can occur in nature and brought me enormous enjoyment. Yet having stood many times at the edges of forests, looking across a bleak landscape of deforestation, I think that perhaps more than anything this work has given me pause to reflect on just how few symbiotic interactions will ever be noticed or understood—and on the countless that have already been lost.

FURTHER READING

THE EVOLUTION AND BIOGEOGRAPHY OF ASSOCIATIONS BETWEEN LYCAENID BUTTERFLIES AND ANTS. N. E. Pierce in *Oxford Surveys in Evolutionary Biology*, Vol. 4. Oxford University Press, 1987.

THE LARVAL ANT-ORGANS OF *THISBE IRENEA* (LEPIDOPTERA: RIODINIDAE) AND THEIR EFFECTS UPON ATTENDING ANTS. P. J. DeVries in *Zoological Journal of the Linnean Society*, Vol. 94, No. 4, pages 379–393; December 1988.

BUTTERFLY EXPLOITATION OF AN ANT-PLANT MUTUALISM: ADDING INSULT TO HERBIVORY. P. J. DeVries and I. Baker in *Journal of the New York Entomological Society*, Vol. 97, No. 3, pages 332–340; July 1989.

HOST SPECIFICITY AMONG *MACULINEA* BUTTERFLIES IN *MYRMICA* ANT NESTS. J. A. Thomas, G. W. Elmes, J. C. Wardlaw and M. Woyciechowski in *Oecologia*, Vol. 79, No. 4, pages 452–457; June 1989.

ENHANCEMENT OF SYMBIOSES BETWEEN BUTTERFLY CATERPILLARS AND ANTS BY VIBRATIONAL COMMUNICATION. P. J. DeVries in *Science*, Vol. 248, No. 4959, pages 1104–1106; June 1, 1990.

THE MUTUALISM BETWEEN *THISBE IRENEA* BUTTERFLIES AND ANTS, AND THE ROLE OF ANT ECOLOGY IN THE EVOLUTION OF LARVAL-ANT ASSOCIATIONS. P. J. DeVries in *Biological Journal of the Linnean Society*, Vol. 43, No. 3, pages 179–195; July 1991.

Diamond Film Semiconductors

Thin sheets of diamond, grown from a low-pressure gas and doped with impurities, may serve as the basis for a new generation of electronic devices

by Michael W. Geis and John C. Angus

Diamond is the hardest known substance and the least compressible. It is a better conductor of heat at room temperature than any other material and, when free of defects and impurities, one of the most transparent. These qualities can be traced to the material's unusual bonding: the atoms of a diamond crystal are more densely packed than those in any other known material. Furthermore, the forces that tie the atoms to one another are stronger than those in most solids.

In addition to their scientific interest, such extreme properties make diamond technologically useful. The gem has found application as an industrial abrasive, as a surgical cutting tool and as a heat sink for cooling electronic components. Researchers have also shown that diamonds containing various impurities function as semiconductors, although natural crystals are too small and expensive for actual use.

That state of affairs is changing. During the past five years, workers have developed a wide range of techniques for laying down diamond films ranging in thickness from more than a millimeter to as little as a few hundred atoms. These techniques, some of which produce large, flat crystals, could make diamond semiconductors practical. The resulting devices would be able to operate many times faster than their silicon counterparts and at temperatures up to 700 degrees Celsius, which would destroy any other electronic device.

MICHAEL W. GEIS and JOHN C. ANGUS study the deposition and characteristics of diamond films. Geis works at the Massachusetts Institute of Technology Lincoln Laboratory, where he has developed diamond transistors and techniques for etching circuit patterns into diamond. He received a Ph.D. in space physics and astronomy from Rice University in 1976. Angus has been working on diamond growth for more than 25 years. His group at Case Western Reserve University synthesized the first semiconducting diamond films in 1971. Angus received his Ph.D. in chemical engineering from the University of Michigan in 1960.

To turn that potential into reality, we and many other researchers are refining growth methods to produce larger, higher-quality crystals. We and other workers are also investigating methods of introducing precisely controlled impurities into diamond to render it conductive and devising ways to etch the material to form circuit patterns.

Ever since the late 18th century, when Antoine Laurent Lavoisier discovered that diamond is a crystalline form of carbon, experimenters have tried to synthesize it. The first to succeed was William G. Eversole of Union Carbide, who, late in 1952, grew diamond seed crystals from carbon monoxide and hydrocarbon gases at moderate pressure. Most of the earliest efforts, however, were directed toward high-pressure processes that could convert graphite into diamond.

Graphite is the most stable form of carbon under normal conditions, but at pressures of approximately 60,000 atmospheres and at temperatures exceeding 1,500 degrees Celsius, diamond is the thermodynamically stable phase. Natural diamonds appear to have crystallized under similar pressures. In high-pressure growth processes, carbon is dissolved in a molten transition metal that serves as both a solvent and a catalyst; over time, diamond crystals nucleate and grow.

A Swedish group at Allmanna Svenska Elektriska Aktiebolaget (ASEA) synthesized diamond at high pressure in 1953, but inexplicably they elected not to make their discovery known. General Electric announced its success in 1955. High-pressure growth processes require very expensive equipment, and only the very largest industrial enterprises have undertaken them. The total production of diamonds is only about 100 tons per year.

During the time that high-pressure diamond growth techniques were being refined, many workers regarded low-pressure diamond synthesis as the equivalent of transmuting lead into gold—a violation of fundamental thermodynamic principles. Graphite is more stable than diamond at low pressures, the reasoning went, and so diamond should not form, or if it did, it would be immediately converted to graphite. In fact, the existence of a more stable form of carbon (graphite) does not preclude the growth of the metastable one (diamond). Metastable materials routinely form in nature. Indeed, the carbon that forms as soot from a candle is also metastable; it is in an even higher energy state than is diamond.

Materials can remain in a metastable state indefinitely as long as the energy barriers that must be surmounted to reach a more stable state are high enough. The difference in energy between the carbon atoms in diamond and those in graphite is only about 0.5 kilocalorie per mole, considerably less than the energy necessary to melt margarine. To transform diamond into graphite, however, requires adding enough energy to break virtually all its bonds—in effect to boil the diamond into vapor. If diamond can once be made to form from a low-pressure gas, it will not readily transform into graphite.

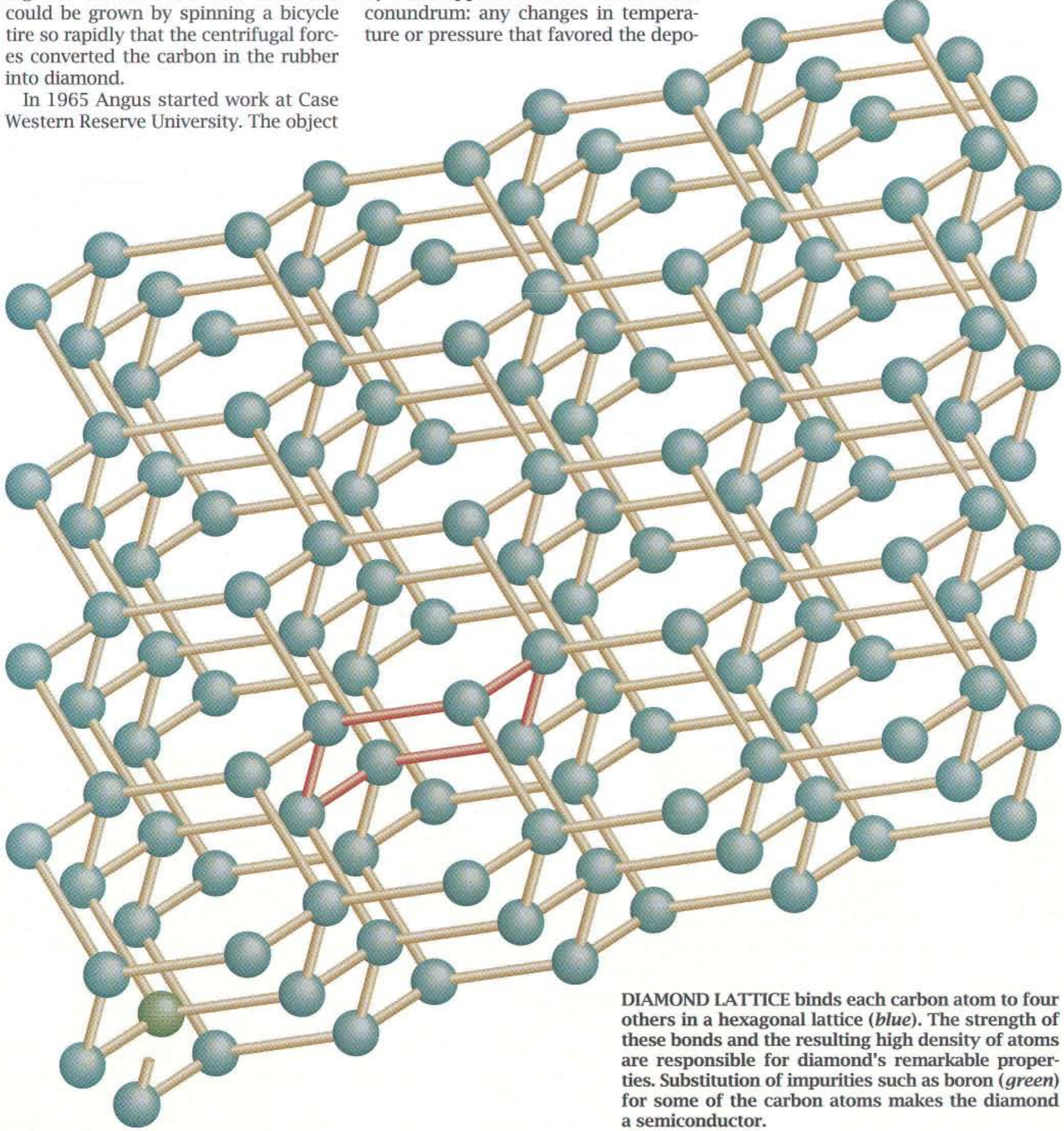
Even though low-pressure diamond growth was clearly not impossible, the few workers who attempted it put their careers in jeopardy. The history of the field is rife with false starts and dead ends. In addition to Eversole's pioneering work at Union Carbide, General Electric also had a low-pressure diamond synthesis program, but both companies stopped work in the area after General Electric announced success with high-pressure synthesis. In 1956 Boris V. Spitsyn, a graduate student at the Physical Chemistry Institute in Moscow, pro-

posed growing diamond by the thermal decomposition of carbon tetraiodide onto diamond seed crystals. And in 1959 one of us (Angus), then a graduate student at the University of Michigan, proposed several similar methods. None of these proposals was made public, and early workers were apparently unaware of one another's endeavors.

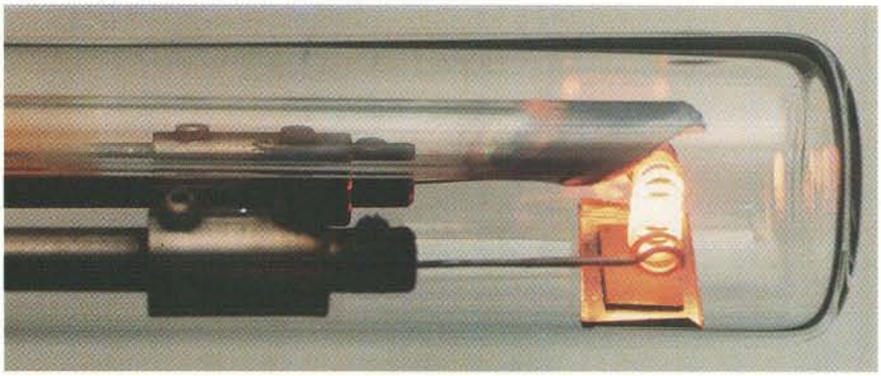
By the mid-1960s the field was moribund. Essentially the only sign of life was a large body of patent literature in which some of the claims bordered on the ludicrous. Perhaps the most amusing was the assertion that diamonds could be grown by spinning a bicycle tire so rapidly that the centrifugal forces converted the carbon in the rubber into diamond.

In 1965 Angus started work at Case Western Reserve University. The object of the research was to obtain convincing experimental evidence of low-pressure diamond growth. Angus and his colleagues reproduced Eversole's results within a year. To prove that what was being synthesized was indeed diamond, they added small quantities of diborane to the methane source gas and grew blue semiconducting diamond. (Other forms of carbon do not show this effect when doped with boron.) The growth rates, however, were too slow to be commercially significant.

These low growth rates were caused by what appeared to be an insoluble conundrum: any changes in temperature or pressure that favored the deposition of diamond favored the deposition of graphite even more. The key to progress has been finding ways to suppress graphite formation, most often by means of atomic hydrogen. Atomic hydrogen reacts with any graphitelike precursors to form molecular structures more closely resembling the structure of diamond. It also reacts with the free dangling bonds on the diamond surface, yielding a configuration of carbon atoms more closely resembling the interior of the crystal lattice. New carbon atoms deposited on the surface are



DIAMOND LATTICE binds each carbon atom to four others in a hexagonal lattice (*blue*). The strength of these bonds and the resulting high density of atoms are responsible for diamond's remarkable properties. Substitution of impurities such as boron (*green*) for some of the carbon atoms makes the diamond a semiconductor.



DEPOSITION CHAMBER for growing diamond films contains a substrate on which the crystals will form, a tungsten filament and a gas-feed tube that supplies hydrogen gas and a simple hydrocarbon such as methane. The chamber is maintained at about one tenth of atmospheric pressure. The filament heats the gases, breaking their molecular bonds. The resulting atomic hydrogen etches away any carbon that is deposited in the form of graphite and allows only diamond crystals to grow. Workers have used microwave discharges and even oxyacetylene torches as alternatives to the tungsten filament.

then more likely to fit into the crystalline arrangement than they are to bond amorphously or to form graphite-like structures.

Angus and his colleagues at Case Western Reserve first used atomic hydrogen in a cyclic fashion. They deposited diamond and graphite, flushed the surface with hydrogen to etch away the graphite and then deposited more diamond and graphite. They described the method, which employed a hot tungsten filament to dissociate hydrogen molecules, at a meeting in Kiev in 1971. Subsequently, Soviet workers, under the direction of Boris V. Deryagin of the Physical Chemistry Institute, developed a continuous process in which the hydrogen removed the graphite as it was formed.

The Soviet work, reported in 1976, was the first instance in which diamond was formed on nondiamond substrates. For the first time, the growth rate of the material was sufficiently high to be of practical interest. Although the Soviets described the diamonds they had produced, they gave no details of the deposition method, and so few outsiders took the Soviet claims seriously. Moreover, about five years earlier the same group had first announced and then retracted the discovery of "polywater"—an ostensibly polymerized form of water whose behavior turned out to be the result of impurities. This unfortunate episode led to a great deal of skepticism about the validity of claims of low-pressure diamond growth from any source.

Between 1981 and 1983, however, a Japanese group at the National Institute for Research in Inorganic Materials in Tsukuba-shi revealed methods for growing diamond films rapidly at subatmospheric pressures. Nobuo Setaka, Yoichiro Sato, Seiichiro Matsumoto and Mutsukazu Kamo and their colleagues used tungsten filaments and microwave discharges to generate the atomic hydrogen required for growth.

These techniques are still the most popular, but direct-current and radio-frequency discharges are now also being employed. In a most noteworthy innovation, Yoichi Hirose of the Nippon Institute of Technology has deposited diamond from the flame of a slightly fuel-rich oxyacetylene torch. The flame produces both atomic hydrogen and the low-molecular-weight hydrocarbons that condense into diamond.

After the Japanese advances, U.S. and European organizations quickly moved back into the field. Benno Lux of the Technical University of Vienna organized the first European efforts in low-pressure diamond growth. In 1984 Thomas R. Anthony and Robert C. DeVries of General Electric reactivated the company's low-pressure diamond work, and Michael Pinneo founded Crystallume in Menlo Park, Calif., the first firm devoted entirely to diamond-film manufacture. By 1985 the U.S. Office of Naval Research was sponsoring substantial work at Pennsylvania State University, North Carolina State University and the Massachusetts Institute of Technology Lincoln Laboratory. Low-pressure diamond synthesis is now being investigated in hundreds of laboratories around the world.

The ability to grow diamond films opens up the possibility of making diamond-based semiconductors, but many obstacles still must be overcome. The first hurdle is the production of films actually suitable for diodes and transistors. The vast majority of growth techniques produce polycrystalline films consisting of myriad tiny grains. The atomic disorder at the grain boundaries makes these films unsuitable for semiconductors.

One technique for making large, high-quality films uses a template: a single-crystal substrate of a material whose crystal structure and atomic spacing approximate that of diamond. This substrate forces the diamond to grow with a fixed crystallographic orientation. Crystals of nickel, copper, silicon carbide, beryllium oxide and boron nitride have suitable lattice spacings and have been used as templates for growing diamond crystals, a process called heteroepitaxy. Thus far, however, diamond crystals grown on substrates of the first four materials are too small and have too many defects to be used for electronic devices. Some of the highest-quality heteroepitaxial diamond has been grown on boron nitride, but single-crystal boron nitride substrates are even more difficult to grow than single-crystal diamond substrates.

Collaborators at Lincoln Laboratory, Wayne State University, the Naval Research Laboratory, North Carolina State and Case Western Reserve have developed an alternative technique. They employ an array of small seed crystals, all of which have the same crystallographic orientation. Diamond deposited on top of the seeds maintains that orientation.

The seeds are made in a few hours by a standard high-pressure growth process. Workers clean the crystallites and then sort them by size and shape. A series of meshes winnows the crystals, leaving only those about one tenth of a millimeter across. The uniformly sized powder is poured onto a sloped, vibrating table. Rounded crystallites roll downward and are discarded, but the vibration drives faceted crystals upward and into a collection slot at the top of the table.

While the seeds are being prepared, workers use standard integrated-circuit etching techniques to form a matrix of pyramidal pits, each also about a tenth of a millimeter across, on the surface of a silicon wafer. Because silicon has the same crystal structure as diamond, the shape of the pits is an exact match for the octahedral diamond crystals.

Workers suspend the seeds in a liquid to make a slurry and immerse the wafer in it. The seed crystals settle out, and some of them lodge in the pits; the rest are washed away. With well-faceted diamonds and careful control of the clean-

liness of the slurry, more than 95 percent of the pits will contain diamonds oriented within several tenths of a degree of one another.

The workers then deposit a 250-micron-thick diamond layer on the array. This step yields a continuous thick film that is self-supporting when the substrate used to hold and orient the diamond seeds is etched away. The film contains low-angle grain boundaries because the seeds are never perfectly aligned, but these boundaries do not appear to compromise its electrical properties. For many purposes, films containing such boundaries are equivalent to single-crystal substrates.

Diamond substrates produced by this technique are among the largest grown to date, up to an inch in diameter. Standard semiconductor fabrication equipment, however, requires circular substrates two inches or more across. Although growing films of that size does not appear to present insuperable difficulties, the experiment has yet to be attempted.

Once diamond films attain the size and quality required for electronic circuits, they must be rendered conductive. In a perfect diamond crystal, each carbon atom is bound to four others by the valence electrons in its outer shell. To liberate one of these electrons so that it is free to conduct electricity requires 5.5 electron volts. This energy requirement is so large that at room temperature perhaps one or two electrons in a one-gram crystal might be mobile; for practical purposes, the material is an insulator.

If one of the carbon atoms in the lattice is replaced by an atom that has five valence electrons, the resulting extra electron is not strongly bound. It can move and so renders the diamond conductive. Conversely, if one of the carbon atoms is replaced by an atom with three valence electrons, a "hole" appears—a mobile, positively charged apparent particle resulting from the absence of an electron where it is supposed to be. Conductivity resulting from extra electrons is called *n*-type because the charge carriers are negative, and conductivity resulting from holes is called *p*-type.

Boron is the logical candidate for *p*-type doping of diamond. It has three electrons in its outer shell, one fewer than the number needed to satisfy carbon's bonding requirements. In addition, boron is immediately adjacent to carbon in the periodic table, and so it is small enough for incorporation in the tight diamond lattice.

Boron doping of diamond, however, is not very efficient. To move a hole from the neighborhood of the boron atom into the diamond lattice where it can carry current requires 0.36 electron volt. This is roughly 10 times the thermal energy of the atoms in the lattice. As a result, less than 1 percent of the boron atoms in diamond contribute holes to the semiconductor. If the boron-doped diamond is heated, thereby increasing the thermal energy, more holes escape from the boron atoms, making the material more conductive.

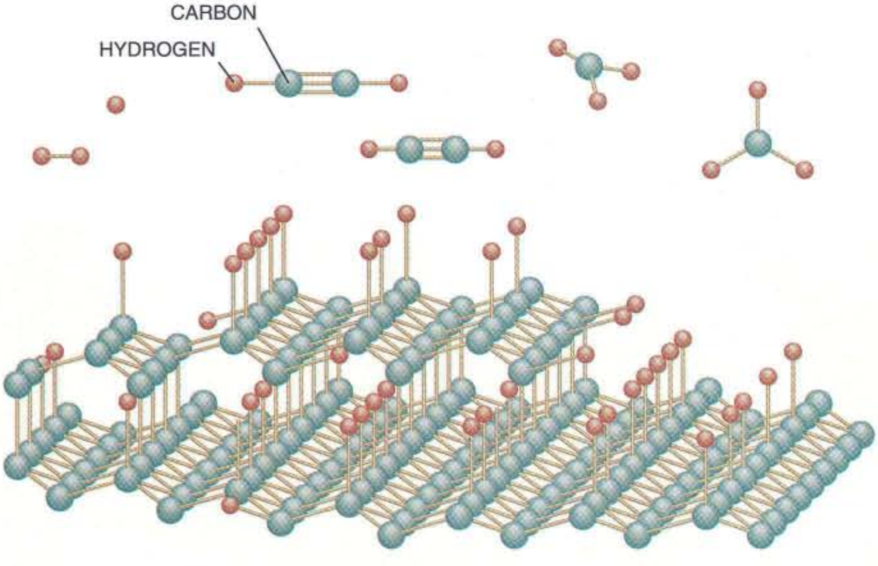
Unlike most semiconductors, diamond can work better at high temperatures: between 100 and 500 degrees Celsius. This behavior is very different from that of silicon, for example, in which boron doping produces weakly bound holes, most of which are free to move at room temperature. Only under the extreme cold of liquid helium will resistivity rise as holes become bound to boron atoms in these materials. (At high temperatures, in contrast, many of silicon's bonding electrons become mobile, making it impossible to control conductivity.)

Although diamond is much more resistive than silicon when it is doped to the same boron impurity level, the close match between the size of the boron and the carbon atoms makes it possible to add far more boron atoms to the lattice than is possible in other semiconductors. This increased doping level compensates somewhat for the high binding energy of individual holes. Indeed, doped diamond can conduct electric current nearly as well as silicon can.

The converse problem, *n*-type doping, is more difficult. Nitrogen would seem the obvious choice because it has five electrons in its outer shell and is adjacent to carbon in the periodic table. Furthermore, nitrogen is a small atom and readily dissolves in diamond. Yet each extra electron is bound to its parent atom with an energy of roughly 1.7 electron volts. At room temperature, only a minuscule fraction of the extra electrons is free to roam the crystal, and so nitrogen is not an efficient *n*-type doping agent.

Phosphorus also has five outer electrons and would seem the next best candidate. Indeed, the limited experimental evidence suggests that adding phosphorus does produce *n*-type semiconducting diamond. Because an atom of phosphorus is significantly larger than one of carbon, however, it is impossible to dissolve any significant amount of it in diamond. In short, *n*-type doping of diamond has thus far proved ineffective.

As a result, most of the diamond-based semiconductor devices that have been made employ positive charge carriers. To fabricate a transistor, for example, one deposits in succession a layer with a very high boron doping level (in which perhaps a few percent of all atoms are boron), a lightly doped layer and another heavily doped layer. The top and bottom layers carry current in or out of the device and generally have low-resistance metal connections attached to them. Current flow through



SURFACE OF GROWING DIAMOND CRYSTAL is almost completely covered by hydrogen atoms (*red*). The bonds between the hydrogen and the carbon stabilize the positions of the surface atoms and so preserve the structure of the diamond lattice. When simple hydrocarbons (*shown above the surface*) condense on the diamond, the carbon atoms form a chemical bond and are incorporated into the diamond.

the device is controlled by varying the voltage on a gate electrode embedded in the lightly doped middle layer.

The thickness and boron concentration of the middle layer determine the maximum operating voltage and current-carrying capacity of the device. The thicker the film and the lower the boron concentration, the higher the operating voltage, but the lower the current-carrying capacity. A device 15 microns thick and one centimeter square, with a boron concentration of 10^{16} atoms per cubic centimeter (about one boron atom per 20 million carbon atoms), can withstand up to 3,000 volts. It has a resistance in the conducting state of about 0.1 ohm; at 225 degrees, its resistance falls to 0.01 ohm. It can carry a current of 10,000 amperes at 3,000 volts.

Although workers can specify precisely both the thickness and the boron level of diamond films, the concentration of other impurities, such as nitrogen, is more difficult to control. Even one nitrogen atom per billion carbon atoms can degrade a film's electrical properties substantially. Yet such small concentrations are almost impossible to detect using standard analytic methods. The development of techniques for ensuring purity is a prime concern for future diamond development, next in importance only to producing large crystals.

Not only is it difficult to control the impurities in diamond films, but it is not feasible to etch them using standard semiconductor manufacturing techniques. Circuit patterns are generally cut into chips by means of a gas that reacts with the substrate to form a volatile compound. For example, oxygen etches carbon-containing compounds such as graphite or plastics by forming gaseous carbon monoxide or carbon dioxide. The oxygen is adsorbed on the surface of a substrate, and then a laser or a beam of atoms that have been accelerated to high speed supplies the additional energy needed to initiate the etching reaction.

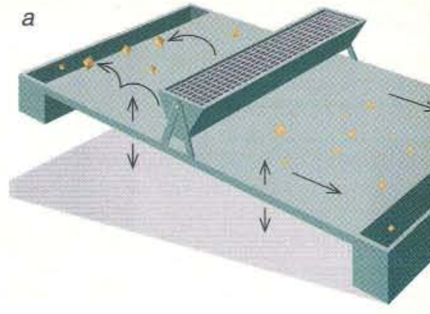
Oxygen attacks diamond several orders of magnitude more slowly than it does graphite and other carbon-containing compounds. The diamond surface is relatively impervious to attack, perhaps because very few oxygen atoms are adsorbed on it. When a beam of fast atoms strikes the surface, there is very little oxygen present to participate in etching.

Instead of oxygen, workers etching diamond employ nitrogen dioxide, which is more easily adsorbed. Etching proceeds at more than 10 times the rate obtained with oxygen alone. Furthermore, the layer of nitrogen dioxide absorbs most of the energy of the incoming atomic beam. Thus, diamond does not suffer the crystal damage that often results from etching with fast atoms.

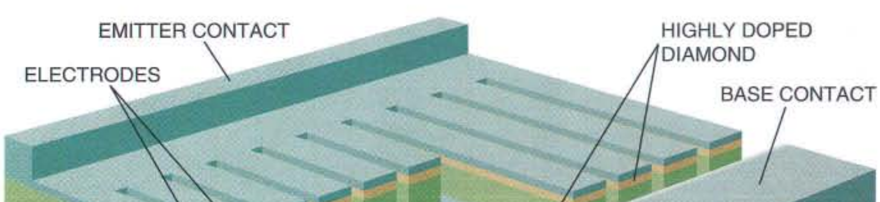
The electrical properties of diamond make it particularly suitable for high-voltage transistors. Diamond breaks down at electric fields between 20 and 50 times those that trigger uncontrolled conduction in silicon or gallium arsenide. The higher breakdown voltage means that the maximum operating voltages of diamond transistors will be proportionately higher than those of similar transistors made of other materials. Moreover, a high-power, high-voltage diamond transistor will be smaller than its silicon or gallium arsenide counterparts of the same rating. The shorter dimensions of the diamond device mean that charge carriers have less distance to travel and are simultaneously subject to a more intense electric field. As a result, it could operate between 40 and 100 times faster than a silicon or gallium arsenide device.

Diamond's 5.5-volt band gap (a measure of the energy required to liberate one of its bonding electrons) makes it ideal in devices that must operate at high temperatures. Diamond Schottky diodes, which consist simply of a junction between metals and diamond, can, for example, operate above 700 degrees Celsius. Such a device could be used in a digital engine controller or similar system operating in a high-temperature environment. The same devices made in silicon, whose band gap is only 1.1 electron volts, do not operate efficiently above 150 degrees.

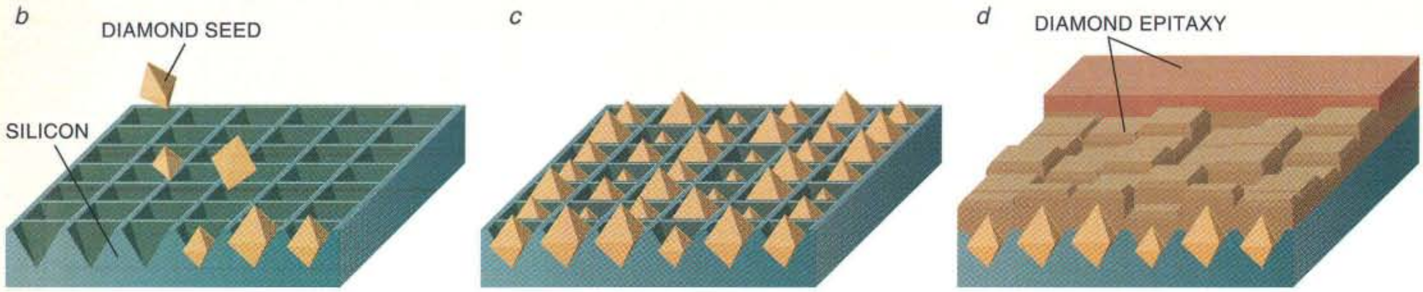
Diamond is also relatively unusual among semiconductors in its ability to form high-quality interfaces with silicon dioxide. In silicon, such interfaces are the key ingredient of the metal-oxide semiconductor field-effect transistor (MOSFET), the most common type of transistor used in computers and other digital applications. A MOSFET consists of a semiconductor coated with a layer of silicon dioxide, on top of which lies a metal electrode. Voltage applied to the electrode produces an electric field that penetrates the insulating silicon dioxide and controls the concentration of mobile carriers (either holes or electrons) below. Because the conductivity of the semiconductor is directly related to the carrier concentration, the voltage



DEPOSITION of large, virtually defect-free diamond films begins with diamond seed crystals. The crystals are sorted by size and shape by means of meshes and a tilted, vibrating table (*a*). Meanwhile



DIAMOND TRANSISTOR consists of three layers (*yellow, green and yellow*), each doped with a different level of boron. The top and bottom layers (emitter and collector) carry current into and out of the transistor. Current through the middle (base) layer is controlled by a voltage applied to electrodes (*blue*) embedded in it.



an array of pyramidal pits is etched into a wafer of silicon (*b*). When a slurry containing the seeds washes over the substrate, the seed crystals settle into the pits with their facets aligned within a fraction of a degree (*c*). Subsequent deposition of diamond produces a film that contains low-angle grain boundaries at the meeting of diamond lattices growing outward from each seed (*d*). These boundaries do not affect the film's electronic properties. Films as wide as an inch across have been grown by this method, and the authors believe larger films are also possible.

on the electrode controls the current flowing through the transistor.

The interface between silicon dioxide and most semiconductors is not suitable for MOSFETs. When a voltage is applied to an electrode on the silicon dioxide, fixed charges quickly accumulate at the interface and cancel out most of the electric field. The field does not penetrate the semiconductor, and so the voltage has little or no effect on the concentration of charge carriers.

Several research groups have now fabricated diamond MOSFETs. The MOSFET is one of the structures most often used in the design of integrated circuits. More important, however, diamond MOSFETs can work much more efficiently than other kinds of diamond transistors. If a negative voltage is applied to the gate electrode, additional holes appear at the interface between the silicon dioxide and the diamond. The current-carrying capacity of the device increases to more than 10 times that achievable by boron doping alone.

Diamond's ability to make high-quality connections to other materials extends to metals as well. When interest was first directed to the development of diamond transistors, researchers feared that diamond, like many other wide-band-gap semiconductors, would make only poor-quality, high-resistance contacts. High-resistance contacts would compromise many of the potential advantages of the diamond transistors by limiting their current-carrying capacity.

The resistance of a contact between diamond and metal decreases as the doping impurity level in the diamond increases. Low-resistance ohmic contacts, which allow unrestricted current flow both in and out of the diamond, can be made simply by using the highest practical doping impurity. Researchers at the Naval Ocean Systems Center and at North Carolina State University have demonstrated that boron concentrations of several percent result in low-resistance contacts comparable to those obtained in silicon.

Although solid-state circuits made of diamond are clearly within reach, the material's ultimate future may be in devices that more closely resemble the vacuum tube than the transistor. In addition to its other useful electrical properties, diamond has a stable, intrinsic negative electron affinity. Conduction electrons can easily escape from a diamond surface into a vacuum, provided that as they leave they are replaced from another source. Expelling electrons from most materials requires temperatures between 1,000 and 3,000 degrees Celsius, or electric fields greater than 10 million volts per centimeter. Semiconductors such as silicon and gallium arsenide can be given a negative electron affinity by coating them with a thin layer of cesium, but extremely small amounts of water or oxygen can react with the coating and destroy its effect.

Once free of the diamond surface, these electrons can be accelerated by an electric field, formed into a beam to strike a target (perhaps a phosphor, which will glow in proportion to the current impinging on it) or modulated by an external signal. So-called cold diamond cathodes could be used in small gas discharge lamps, flat-panel computer and television displays, and perhaps even submicron triodes that would outperform conventional semiconductors.

One of us (Geis) recently fabricated such cathodes. The prototypes are not adversely affected by exposure to water, oxygen or air. They operate at an estimated current density of about 10 amperes per square centimeter, compared with about 50 amperes per square centimeter for the hot cathodes employed in vacuum devices such as cathode-ray tubes or high-power amplifiers for radio and television transmitters.

Because neither heat nor excessive voltage need be applied to force electrons out of diamond cathodes, vacuum tubes containing them should not be subject to the factors that limit the life of conventional cathodes. Indeed, the current-carrying capacity of a diamond cathode is limited for practical purposes only by the resistance of the conducting substrate that supplies it with electrons. In theory, these devices could operate reliably at current densities of more than 1,000 amperes per square centimeter.

At such current levels, they could be the basis of subminiature vacuum tubes that would have higher power and frequency capabilities than do even diamond transistors. Paradoxically, the highly refined product of solid-state electronics would lead to a resurgence of the vacuum tubes that transistors eclipsed more than 30 years ago.

FURTHER READING

DIAMONDS. Eric Bruton. Radnor, Pa., Chilton Book Company, 1978.

LOW-PRESSURE, METASTABLE GROWTH OF DIAMOND AND "DIAMONDLIKE" PHASES. John C. Angus and Cliff C. Hayman in *Science*, Vol. 241, No. 4868, pages 913–921; August 19, 1988.

APPLICATIONS OF DIAMOND FILMS AND RELATED MATERIALS. Edited by Y. Tzeng, M. Yoshikawa, M. Murakawa and A. Feldman. Elsevier, 1991.

DIAMOND CHEMICAL VAPOR DEPOSITION. F. G. Celii and J. E. Butler in *Annual Review of Physical Chemistry*, Vol. 42. Annual Reviews, Inc., 1991.

LARGE BANDGAP ELECTRONIC MATERIALS AND COMPONENTS. Special issue of *Proceedings of the IEEE*, Vol. 79, No. 5; 1991.

METASTABLE GROWTH OF DIAMOND AND DIAMOND-LIKE PHASES. J. C. Angus, Y. Wang and M. Sunkara in *Annual Review of Materials Science*, Vol. 21. Annual Reviews, Inc., 1991.

Scavenging and Human Evolution

Although meat eating helped to shape the evolution of human brains, behavior and toolmaking, our early ancestors seem to have been better scavengers than hunters

by Robert J. Blumenschine and John A. Cavallo

Man the Hunter is a phrase that rings. Who would not rather be numbered with the lion than with the vulture? Hunting seems nobler than scavenging and, at first glance, more profitable, too. What better way to reaffirm our evolutionary success than to portray our earliest hominid ancestors as mighty hunters? Many anthropologists agree that eating the meat of large animals helped to form the physical and social environment that selected for the traits that most distinguish humans from apes. But was that meat acquired by predation or by scavenging? This question matters perhaps as much as any in evolutionary studies because it touches on the definition of human nature. Unfortunately, the answer given by the theory of Man the Hunter is based more on sexual and other prejudices than on the fossil record and the ecology of finding food.

Scavenging has received little notice, we believe, because many anthropologists have been too quick to project current ways of life into the past. They use hunter-gatherers, apes or carnivores as surrogates for aspects of early hominid life that have been obscured

SCAVENGING OPPORTUNITIES differed by size, terrain, season and cause of death. Riparian woodlands probably offered the best returns because the trees provided hominids with refuge and hid carcasses from vultures. Arboreal leopard kills (*1*) were probably available year-round; lion kills (*2*) came mainly in the dry season. Saber-toothed cats seem to have

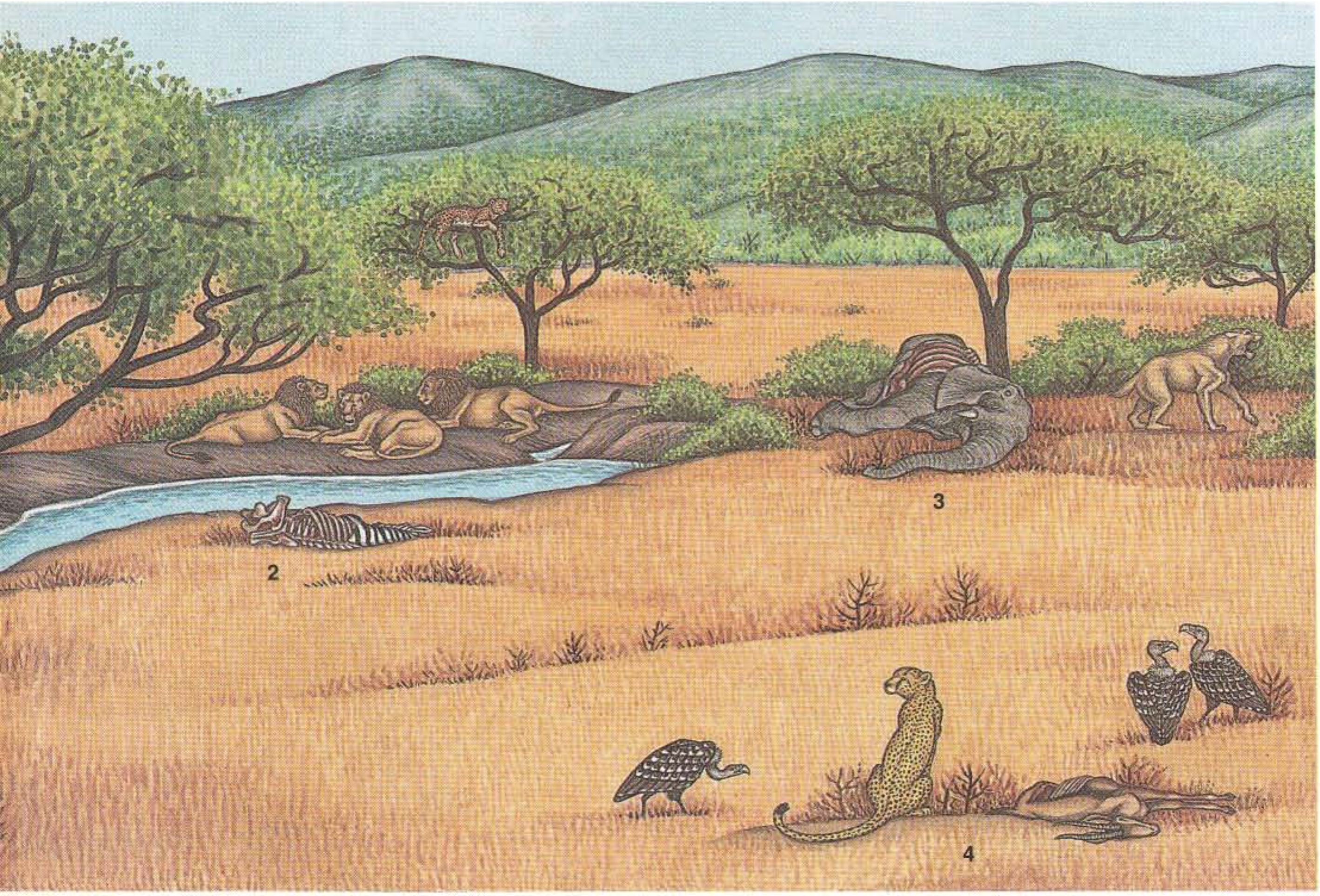
by the passage of time—a practice that strips the hominids of the very adaptations that made them unique. Advocates of the hunting theory also elevate hominids above other organisms, as if our ancestors were immune to most of the pressures that shape relations between predators and prey. In all these matters, they assume that early hominids found hunting to be bountiful, predictable and safe and scavenging to be marginal, opportunistic and risky.

ROBERT J. BLUMENSCHINE and JOHN A. CAVALLO collaborate in the study of early hominids at Rutgers University. Blumenschine, an associate professor of anthropology, specializes in the archaeology of human origins, particularly diet and foraging behavior. He co-directs archaeological research at Olduvai Gorge, Tanzania, and has also done archaeological and related wildlife research in Ethiopia and India. Cavallo, a doctoral candidate in anthropology, directs the Rutgers Center for Public Archaeology. He has studied leopard behavior in Serengeti National Park, Tanzania, and conducted research into prehistoric Native American hunting and fishing economies.

Our research reaches quite different conclusions. Scavenging may have been more common than hunting two million years ago, at the boundary between the Pliocene and Pleistocene epochs. Flaked-stone toolmaking, the practice of butchering large animals and the evolution of big-brained *Homo* all make their first known appearance in the physical record at this time. Because much of the evidence lies at such east African sites as Olduvai Gorge in Tanzania, we attempted to learn how to decipher the residues of ancient subsistence patterns at nearby game reserves: Tanzania's Serengeti National Park and Ngorongoro Conservation Area. We also tried to test objectively the prevailing notion that scavenging would have been inferior to hunting.

In independent stints over a period of 20 months, we noted how predators and scavengers got their meat and what they did to the bones they left behind. Our fieldwork thus united ethology with taphonomy—the study of how postmortem events alter carcasses in the fossil record. Further, we integrated these results with paleontological and archaeological evidence for the behavior of protohominids. This approach reads into the past only those aspects of present behavior and ecology that leave preservable residues. It thus avoids a wholesale imposition of the way of life of a modern species that happens to suit one's ideals.

That no substitute exists in anthropology for such actualistic studies may be demonstrated by what zoologists have shown about the behavior of the hyena—the popular symbol of scavenging—and the lion—the prototypical predator. Until 30 years ago, no one conceived that each carnivore both hunts and scavenges. If biases can so cloud



left large carcasses (*3*) in these habitats in all seasons. Open-country kills of lions, cheetahs (*4*) and hyenas (*5*) were less attractive to hominids, who lagged behind the vultures and hyenas and had no place to escape from larger carnivores. Drownings (*6*) and starvation (*7*) afforded windfall opportunities, not all of which were safe to exploit.

SCAVENGING ON THE OPEN PLAIN requires speed and strength beyond what early hominids possessed. These spotted hyenas reached this lion kill before vultures stripped it and could outrun any lion that might return.

the truth about living carnivores, how much more careful must scientists be in reconstructing the subsistence of extinct hominids?

The theory of Man the Hunter has never been constrained by fossil evidence. Charles Darwin was the first to present hunting as the behavioral catalyst that selected for an enlarged brain, tool use, reduced canine teeth and bipedalism, thus splitting the lineages of humans and apes. He laid out his hypothesis in *The Descent of Man* (1871), before any fossils earlier than the Neanderthals had been found. When more ancient specimens turned up in the early decades of this century, workers linked them directly to Darwin's scheme. Raymond A. Dart, discoverer of the *Australopithecus* genus, spent some 30 years trying to show that this hominid could have hunted the animals whose bones were so often found mingled with its own. To circumvent the problem of the absence of stone tools at these sites, Dart invoked an "osteodontokeratic" tool and weapon kit made from animal bones, teeth and horn.

This interpretation gained popular support in the many accounts of humanity's "killer ape" forebears. It fell apart, however, under the critical tests of the pioneering taphonomist C. K. Brain of South Africa's Transvaal Museum. He showed that the australopithecines had played no role in gathering the bones of the animals found in association with their own skeletons. Instead, these studies suggested, both hominids and ungulates had ended together when the leopards that hunted them discarded their carcasses at the base of their favored feeding trees. Yet the hunting hypothesis remained intact; now, however, it was made to apply to the later stage in evolutionary history that began with the appearance of large-brained *Homo habilis.*

The arguments for this theory reached full flower in the papers collected in Richard B. Lee and Irven DeVore's *Man the Hunter* (1968). The contributors sketched the following scenario. Protohominids encroach on the savanna by eking out their accustomed vegetarian diet with increasing amounts of hunted flesh. Hunting puts a premium on foresight and dexterity, selecting for larger brains and nimbler hands. These traits increase the capacity for technology, raising the payoff of intelligence and augmenting the original selective pressure. Hunting becomes the engine of a self-sustaining cycle of social and intellectual evolution.

This theory prevailed until the late 1970s, when an influential article by the late Glynn Isaac shifted the emphasis from the gathering of meat to the sharing of it [see "The Food-Sharing Behavior of Protohuman Hominids," by Glynn Isaac; SCIENTIFIC AMERICAN, April 1978]. Isaac, an archaeologist at the University of California at Berkeley, showed that early hominids had home bases—a behavioral innovation—which, he argued, implied a sexual division of labor—another innovation. To enhance the omnivorous strategy, males ranged far in search of scavengeable meat or hunted quarry, females gathered fruits and tubers nearer home and families shared the take. Eventually this altruistic behavior and social cooperation began to select for intelligence, language and culture.

Lewis R. Binford, now at Southern Methodist University, carried such analysis further in 1981. In a taphonomic reanalysis of Mary Leakey's data from the early Olduvai bone assemblages, Binford argued that neither hunting nor food sharing had evolved by *H. habilis* times. Hominids had merely processed the meager leftovers of more capable carnivores by breaking open bones to get at the marrow. He said scavenging could not have provided the surpluses of meat needed to sustain food sharing. Instead the social and nutritional aspects of protohominid feeding resembled the mainly vegetarian diets of modern apes.

Binford later argued, on similar grounds, that even the early modern *Homo sapiens* of southern Africa and the contemporaneous Neanderthals of Europe relied on scavenging to get large animals and hunted only small ones. Thus, Binford, too, retained the hunting hypothesis by moving it closer to the present—within the past 100,000 years. His reconstruction accepts that scavenging was a penurious enterprise and that hunting and its attendant pattern of food sharing was a driving evolutionary force, albeit one that took effect very late in our evolution.

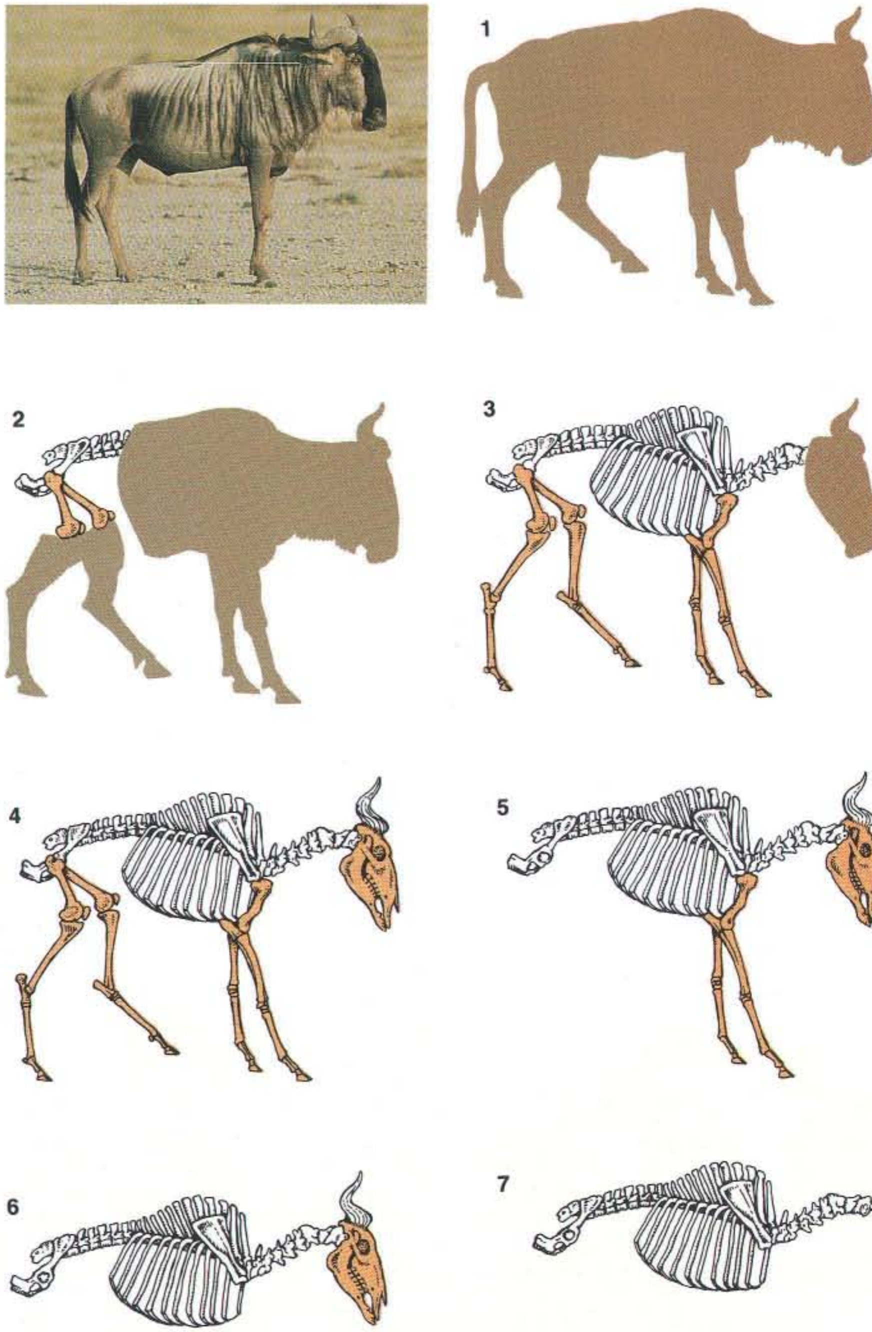
We began our critique of this entire approach by appraising the hunting prowess of early hominids. The physiques of *Australopithecus* and early *Homo* were unprepossessing. Females stood about four feet tall, males under five; females weighed about 70 pounds and males around 100. Their long arms suggest they still took refuge in trees. No doubt they had frequent occasion to do so, confronted as they were by such proficient predators as lions, saber-toothed cats and hyenas. As for their tools, even *Homo* wielded a very primitive kit of rough-hewn scrapers and unworked hammerstones. No true weapons are apparent.

Yet the archaeological evidence shows that these puny primates encroached on the large carnivores' niche. At Olduvai and elsewhere, archaeologists have found simple stone artifacts in association with fossilized bone fragments from animals as small as a gazelle and as big as an elephant. Surfaces of some of these bones bear the tooth marks of carnivores. Some of these and other bones also carry cut marks made when the associated tools were used to remove meat and disarticulate bones. Many bones are fractured and marked by a hammerstone, used to get at the marrow. Could protohumans have killed animals as fleet and formidable as these? We think scavenging deserves a closer look.

Proponents of hunting have argued that a diurnal hominid would have had difficulty in locating the kills of wider-ranging predators and that any they might have chanced on would have been thoroughly eaten by hyenas, the only animal that can crush bones for marrow with its teeth. But these arguments miss two scavenging opportunities that we identified in Tanzania—big-cat kills in riparian woodlands and the carcasses of very large animals that die of disease or by drowning. Hominids foraging in this habitat may have pieced together a niche no other scavenger could exploit so well [*see illustration on pages 70 and 71*].

Riparian woodlands would have suited partially arboreal bipeds by providing sanctuary and by hiding carcasses from vultures, the lead spotters of the scavenger tribe. Large ungulate carcasses crop up in these regions mainly in the dry season, when lions abandon their defleshed, zebra-size kills. Leopard kills, on the other hand, consist of smaller ungulates and are available year-round. These kills are shielded best of all because they are typically stored in trees. Two million years ago saber-toothed cats may have provided hominids with a third opportunity, also in riparian woodlands. Kills of these extinct predators would have provided very large carcasses and abundant meat.

We hypothesize that scavenging may have been most important in the dry season, when plant foods are scarcest and scavenging opportunities are most diverse. Aside from leopard kills, wet-season predation does not have the predictable riparian focus but instead is scattered in broader and more open habitats. Hyenas are quick to find and consume these exposed carcasses. Because scavenging may have made carnivory and herbivory seasonally complementary feeding strategies, we do not assume—as proponents of Man the Hunter do—that getting meat was the



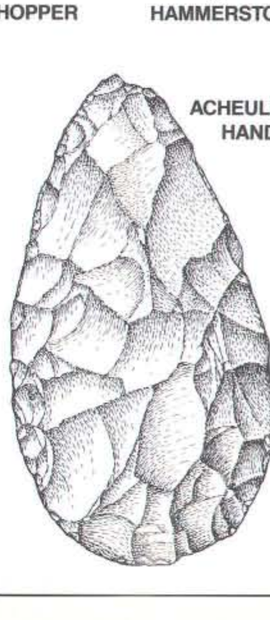
VANISHING WILDEBEEST illustrates the seven stages of scavenging. Only hyenas and tool-using hominids can exploit carcasses of this size beyond stage four by breaking limb bones for the marrow and opening skulls for the brains. In his study of *FLK Zinjanthropus*, an east African archaeological site dated to about two million years ago, Henry T. Bunn of the University of Wisconsin at Madison found a preponderance of head and limb bones. The finding matches stage 4, suggesting that the hominids scavenged defleshed carcasses.

Primeval Cutlery

Butchering implements included flakes to strip tissues (*top*) and hammerstones to break bones. Such tools are classed as Oldowan because they appear in the early levels at Olduvai Gorge. The succeeding Acheulean technology is exemplified by a very refined hand ax (*bottom*).

SIMPLE FLAKES

CHOPPER HAMMERSTONE



ACHEULEAN HAND AX

core of hominid adaptation. As dental evidence suggests, hominids have always been omnivores. The mere existence of stone tools and animal bones does not demonstrate that meat eating was common.

Yet scavenging may have made the dry season a time of plenty. It is then that starvation and predation produce many carcasses. Even the most marginal abandoned lion kill, retaining only marrow and brain, could provide much more than an adult's daily caloric requirements at the cost of half an hour's effort with a hammerstone. This rate of food return is higher than can be obtained by harvesting plants. If efficiency guided foraging decisions, hominids would have always preferred scavenging to harvesting, whenever scavenging was possible.

This preference would have been most marked at the height of the dry season, when plant productivity reaches its nadir and big-cat kills become predictable, shortening searches for these resources. A similar economy applies in the comparison with hunting: less energy is spent in picking one's food out of a tree or up off the ground if one does not first have to chase it.

Scavenging also incurs less risk than hunting. Any meat that attracts hominids can also attract lions, which, when they arrive, may ignore the dead in order to pursue the living. Our research showed, however, that large carnivores often leave certain classes of carcasses unattended for long periods. In the interim, the sites would have been safe.

Defleshed lion kills in riparian woodlands are particularly safe. We found that bone-crushing hyenas usually do not discover these carcasses until a day after they are abandoned by lions—a good window of opportunity for any hominid capable of wielding a hammerstone. Tree-stored leopard kills provide more food (flesh as well as marrow) at less risk, especially when the cache contains several kills. Leopards tend to be solitary, and even a baboon or chimpanzee can sometimes scare one away. Moreover, leopards often abandon their kills voluntarily for as long as eight to 12 hours during the day, leaving some of them in a complete condition. Hominids in these woodlands would appear to have faced no more danger in scavenging than in foraging for plant foods in the same regions.

But risks may have outweighed benefits in the open plains, despite the plentiful opportunities for displacing the timid cheetah and jackal from their kills, exploiting abandoned lion kills in wet periods and benefiting from natural deaths during drought. The reason for this judgment is the scarcity of trees, which deprived the arboreally adapted hominids of sanctuary. Yet this drawback applies with at least equal force to hunting. Big herbivores have ways of defending themselves, and even if one can be killed, the conspicuousness of the killing quickly attracts scavengers. Many of these would have been more than a match for rock-wielding bipeds.

Hunting enthusiasts might respond that game is more wholesome than carrion. In the Serengeti, however, we found that few carcasses left on the ground retain scavengeable food as long as 48 hours, the time it takes for putrefaction to set in. But even then, most edible tissues remain encased in skin or bone that excludes insects and other postmortem disease vectors. Even carcasses produced by "natural" death generally carry no dangerous parasites, because most such deaths result from malnutrition, not disease.

Scavenging has also been faulted as nutritionally unsound. John D. Speth of the University of Michigan has suggested that animals that died of hunger would provide protein without enough fat for balance, a diet that can lead to a form of starvation. (Backwoodsmen called it "rabbit fever" because it came from living exclusively on rabbits and other lean game.) Yet hominids have always gotten most of their calories from the carbohydrates and oils of plants, and the most regular dry-season scavenging possibilities are predator kills of animals with fat in their marrow.

Which came first, scavenging or hunting? Answers have been proffered on ethological grounds, only to be controverted by new evidence. Hunting was uniquely human until Jane Goodall documented it in chimpanzees. Scavenging was beneath the dignity of a primate until workers discovered chimpanzees and baboons usurping the kills of cheetahs and leopards. It was foreign to human nature until 1988, when an ethnographic study of 20 years' duration documented avid scavenging by the Hadza and San foragers of sub-Saharan Africa. The delayed observation testifies to the prejudice against scavenging.

The earliest hominids probably scavenged and took small prey with their hands, as chimpanzees and baboons do. Only their next step was unique: they began to use tools to butcher large carcasses that nonhuman primates cannot exploit. The difficulty of this leap belies the charge that scavenging offers no challenge that might select for human qualities.

Our fieldwork suggests that scavenging is not at all easy for a slow, small, dull-toothed primate. To locate scavengeable carcasses before others did, we had to learn how to interpret the diverse cues to the presence of a carcass in riparian woodlands. They include the labored, low-level, early-morning, beeline flight of a single vulture toward a kill; vultures perched in mid-canopy rather than at the crown of a tree, where they nest; appendages of a concealed leopard or of its kill dangling from a branch; and tufts of ungulate hair or fresh claw marks at the base of a leopard's feeding tree. At night, the loud "laughing" of hyenas at a fresh kill, the panicked braying of a zebra being attacked, the grunting of a frightened wildebeest—all serve notice of where to find an abandoned carcass when morning comes.

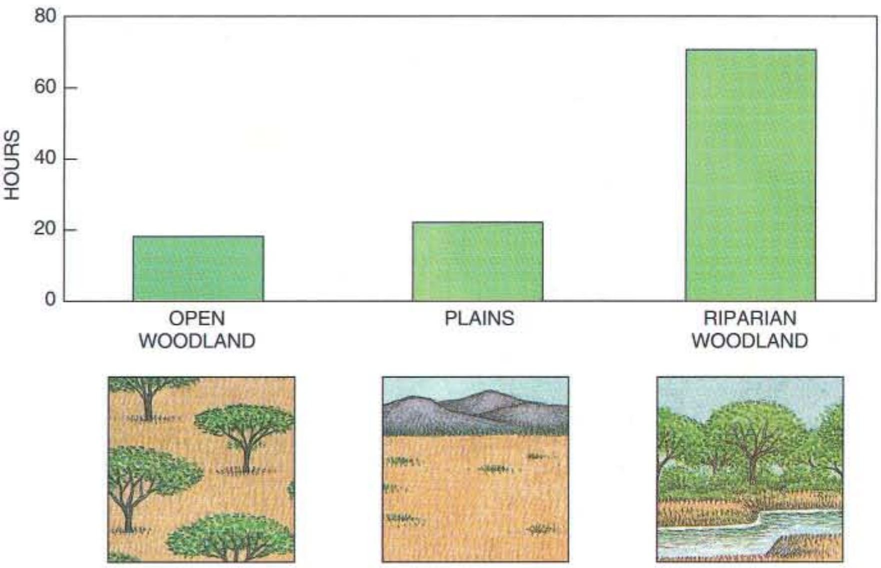
Higher primates make "mental maps" of their ranges and use them to predict where the next batch of fruit will ripen. Hominids might have applied this ready-made skill to predict the future availability and location of carcasses. We learned how to do it, with great effort. Every day we monitored the movements, hunting and feeding schedules, and belly sizes of predators, as well as the general activity of their prey. Apart from its possible nutritional payoffs, hominids might have used such information routinely to avoid predators.

Social skills would not have advanced, however, unless scavenging also selected for social cooperation. Scavenged carcasses that fed only one individual, leaving no surplus to share, would probably have promoted competition. But if our research results are correct and big-cat kills gave early hominids a food surplus, then Isaac's model of cooperative foraging, processing and food sharing would work. Similarly, if such carcass foods did not usually coincide with plant foods, the emergent social skills might have expanded to include a division of labor, with corporate foraging about a common home base. To add to our ancestors' challenges, one need only hypothesize that they generally found carcasses in one place and stone for butchery tools in another. Uniting the tools with their objects would have thus required deep planning depth, detailed mental mapping and social cooperation.

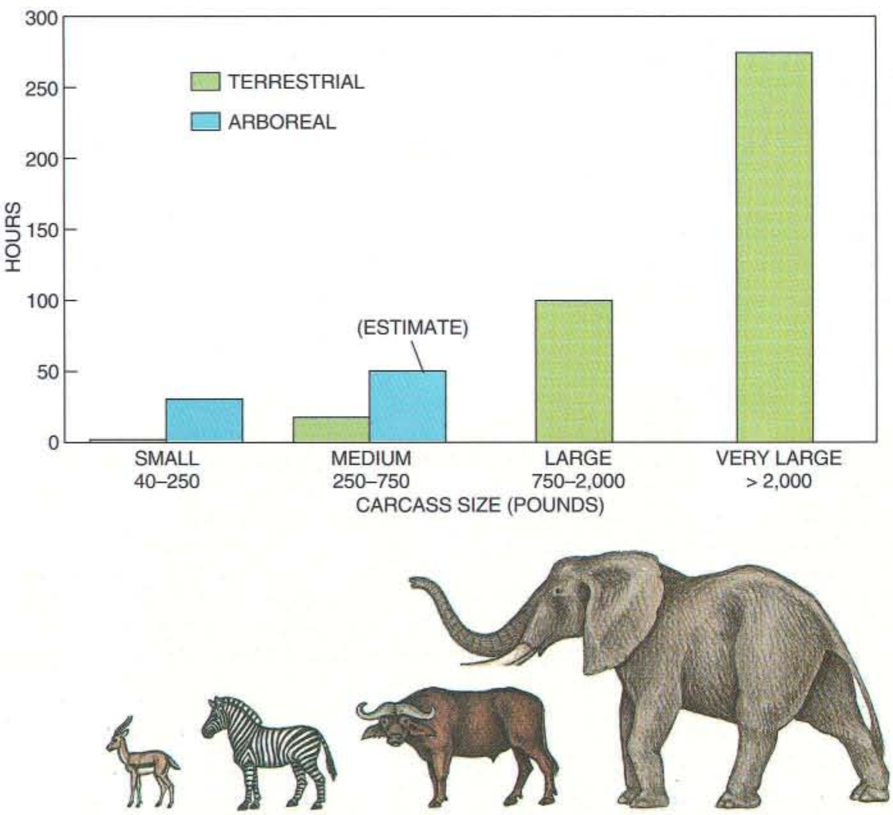
West African chimpanzees are the only nonhuman primates that have enough planning depth to bring stone tools to food sources, as they do when they transport stone hammers and anvils to break the hard nuts of kola and Panda trees. Still, they do not carry the stone very far—*H. habilis* transported stone as far as 10 kilometers (six miles)—and the nuts are not nearly as ephemeral as the scavengeable carcasses the early hominids butchered.

Technological skills necessary for exploiting most scavenging options are embodied in the earliest, Oldowan, tool kit [*see box on opposite page*]—sharp-edged stone flakes to deflesh and disarticulate and natural cobbles to break marrow bones and skulls. No tools clearly designed as weapons are apparent in either this complex or those of the more

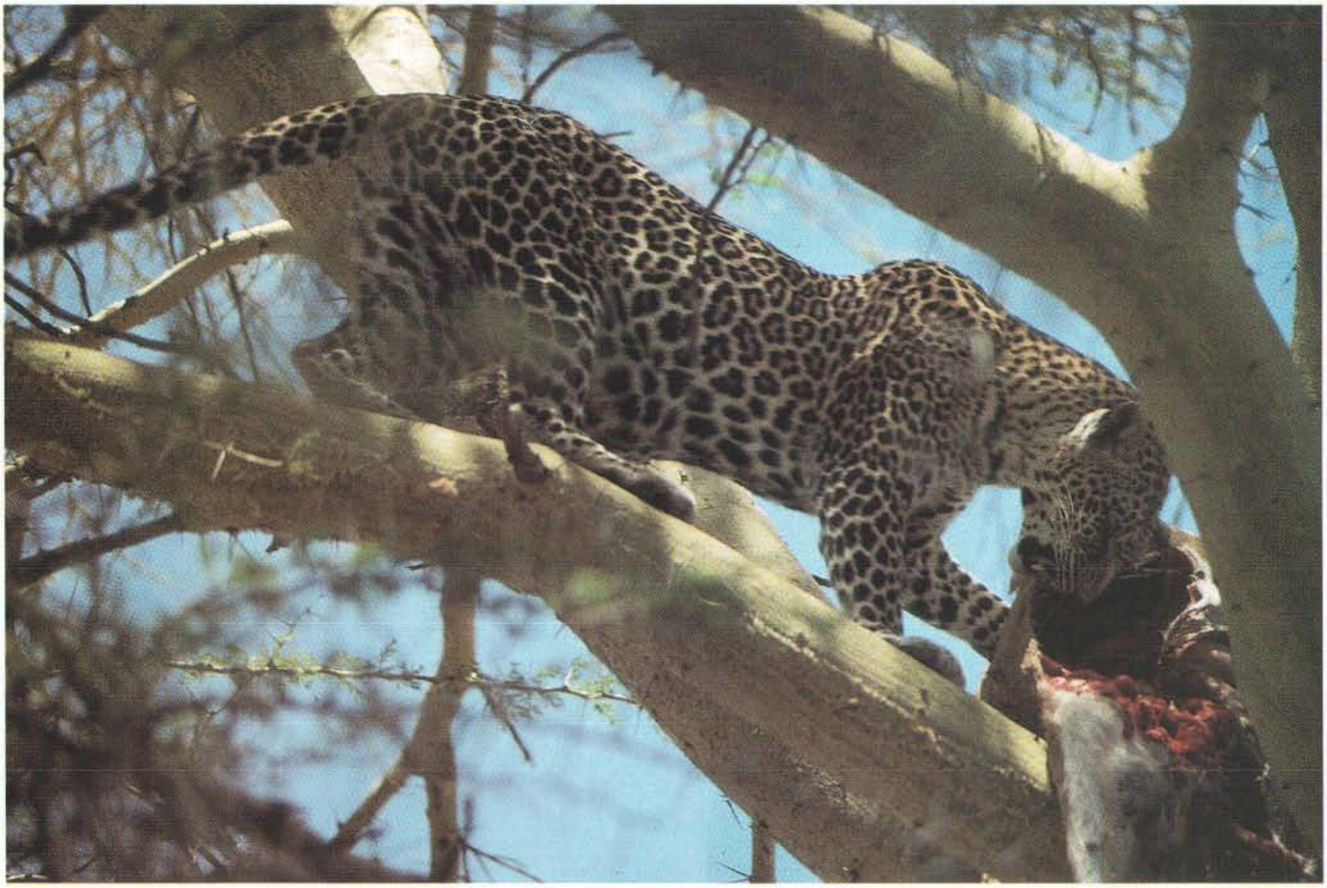
PERSISTENCE OF CARCASSES, BY HABITAT (LION KILLS ONLY)



PERSISTENCE OF CARCASSES, BY SIZE



CARCASS PERSISTENCE in the Serengeti varies by habitat and size. Large carcasses on the ground outlast small ones, tree-stored kills outlast terrestrial ones and carcasses in dense riparian woodlands outlast those in open woodlands and savannas.



LEOPARD'S LARDER keeps carcasses out of reach and out of sight to all except hominids. Because such tree-stored kills may retain flesh as well as marrow, they would have been particularly valuable to hominids with butchering tools.

sophisticated tools of the Acheulean age, which ranged from 1.5 million to 200,000 years ago.

Such considerations lead us to conclude that Oldowan hominids may have created a scavenging niche that can account both for the earliest assemblages of tools and large mammal bones and for many uniquely human traits assumed to have arisen from hunting. A scavenging option may have started as a supplement to plant foraging, appearing in the following stages.

Hominids may have begun eating large animals long before *Homo* appeared. *Australopithecus* could have pioneered the strategy when it occupied the savannas and woodlands that became widespread by six million years ago, in the wake of global climatic change. These open environments would have offered far better scavenging opportunities than the closed woodlands and forests of hominid predecessors—habitats in which the apes have remained to this day.

The earliest hominids may have come across defleshed kills while foraging for plants in thin ribbons of riparian woodlands. These resources would require only unmodified cobbles to extract the remaining marrow and brain from their bony cases. The butchery of these rich sources of energy and protein might have left a record that has so far eluded archaeologists because it occurred before the invention of flake tools, whose manufacture leaves a conspicuous litter of stone chips.

If so, diurnal hominids may have started to supplant hyenas by getting to the kill first. This hypothesis finds some confirmation in the extinction of several hyena species about two million years ago. The advent of a flaked stone technology by about 2.5 million years ago may then have enabled hominids to encroach on a new component of the large mammal scavenging niche. Now they could get flesh as well as marrow. With flaked stone, hominids held in their hands fabricated versions of the flesh-shearing carnassial teeth of carnivores. With them, the hominids could butcher a leopard's tree-stored kills. The flesh from the much larger prey of the saber-toothed cats would also have been accessible, an observation that leads us to suggest that hominids may have had something to do with this species' extinction some 1.5 million years ago. It may be significant that these big cats persisted longer in Europe and the Americas than in Africa, going extinct only after hominids first colonized these continents.

Hunting of very small prey by hominids may have been an ancient strategy, and the late development of projectile weapons made early *H. sapiens* a predator more capable than any other primate. But scavenging has probably had a much more pervasive effect on human evolution than has hitherto been appreciated.

FURTHER READING

THE SERENGETI LION: A STUDY OF PREDATOR-PREY RELATIONS. George B. Schaller. University of Chicago Press, 1976.

SERENGETI: DYNAMICS OF AN ECOSYSTEM. A. R. Sinclair and M. Norton-Griffiths. University of Chicago Press, 1984.

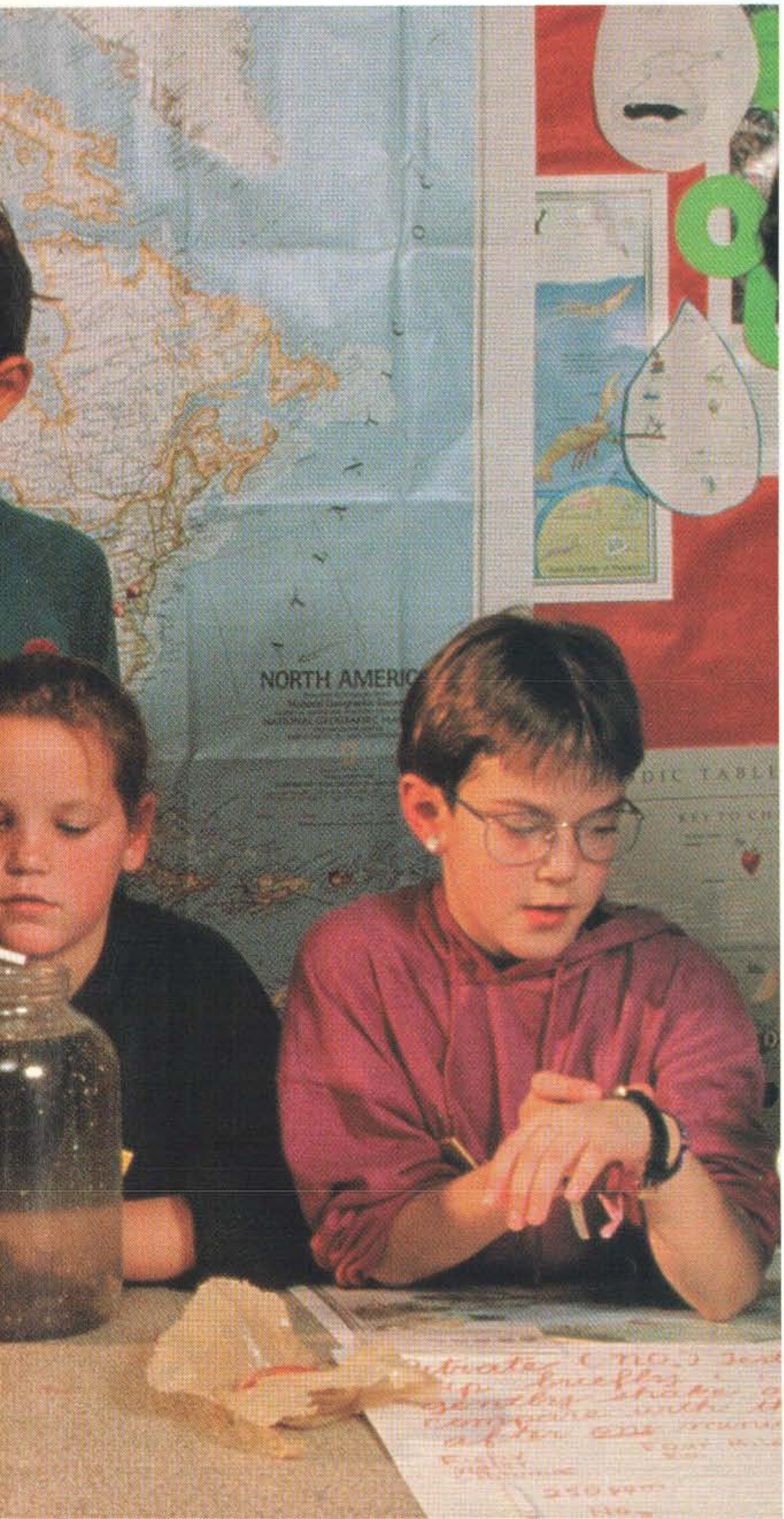
LAST DAYS IN EDEN. Elspeth Huxley and Hugo Van Lawick. Amaryllis Press, 1984.

ANOTHER UNIQUE SPECIES: PATTERNS IN HUMAN EVOLUTIONARY ECOLOGY. Robert Foley. John Wiley and Sons, Inc., 1987.

THE FIRST TECHNOLOGY. Nicholas Toth in *Scientific American*, Vol. 256, No. 4, pages 104–113; April 1987.

TEACHING REAL SCIENCE

by Tim Beardsley, *staff writer*



Educators are planning a revolution in science teaching. They face an uphill struggle.

This past June, a select group of 20 teachers and academics cloistered themselves in the National Research Council's Arnold and Mabel Beckman Center in Irvine, Calif. After two intensive two-week spells of drafting and editing, they produced proposals for a change in U.S. education that would have been unthinkable just a few years ago: voluntary national standards for science curriculums from kindergarten through grade 12. Initial accounts of the work in progress are to be presented at conferences as early as this fall.

The effort has grown out of a widespread conviction that precollege mathematics and science education in the U.S. is in such a grim state that radical reforms are urgent. The Carnegie Commission on Science, Technology and Government, for example, said last year that the failings are "a chronic and serious threat to our nation's future." The commission's report observed that when 47 percent of the nation's 17-year-olds cannot convert nine parts out of 100 to a percentage and when 63 percent of American adults think lasers work by focusing sound waves, "we know that science education in this country is not working."

Teachers of all subjects complain that classes are too large and that they have too little time to prepare lessons. But science is hit particularly hard by inadequate teacher training and by shortages of funds to buy essential and increasingly costly supplies. In addition, science teachers have to contend with uninspiring, error-filled textbooks that are often

SCIENCE IS FUN when youngsters do experiments. Those here will use a computer linked to the National Geographic Kids Network to share findings with others around the world.

designed with an eye more to the religious sensitivities of adoption committees in key markets such as Texas and California than to accuracy. All these factors have conspired to produce a system that actively discourages youngsters from pursuing technical subjects.

Worse yet, aversion to science is not merely socially acceptable among high school students, it is positively hip. The National Assessment of Educational Progress, a congressionally mandated study, reported this year in "The 1990 Science Report Card" that the proportion of students answering "No" to the question "Do you like science?" actually increases between grades 4 and 12, from 20 to 35 percent. "We are violating the Hippocratic oath to first do no harm," says Douglas A. Lapp, director of the National Science Resource Center in Washington, D.C., which promotes educational reforms.

The academics, scientists and teachers who met in Irvine hope the consensus they are seeking on curriculum standards for science will by 1996 provide a yardstick to accomplish what years of exhortation have failed to do. Their unstated goal is to convince school districts to do nothing less than throw out most of their science courses and start over. The curriculum standards are to be "narrative descriptions of what every student should understand about science and its applications," according to the National Research Council (NRC), which coordinates the effort. The council is also overseeing the development of standards for teaching and for assessment. The teaching standards will describe "what teachers should know and be able to do," says Kenneth M. Hoffman, an NRC official who has long experience with science education. The assessment standards will offer criteria for evaluating students and courses.

Educators across the country, anticipating the arrival of the standards, are already striving to produce teaching materials that reflect both modern science and current ideas about learning. They are also, crucially, trying to provide teachers with adequate training to use the materials.

Their work is being conducted against a backdrop of repeated failures. When *Sputnik* raced into orbit in October 1957, its bleeping radio a beacon of Soviet technological prowess, the U.S. responded by boosting precollege science education. But the efforts fell into disrepute. In the early 1970s the political right in Congress attacked a course on social science published by the National Science Foundation (NSF) on the grounds that it offended traditional values and represented federal interference in states' rights. The foundation's science education workshops for teachers were discontinued soon afterward, when studies failed to show they were having any lasting effect. By 1982 the NSF's education programs had all but petered out.

The Second Wave

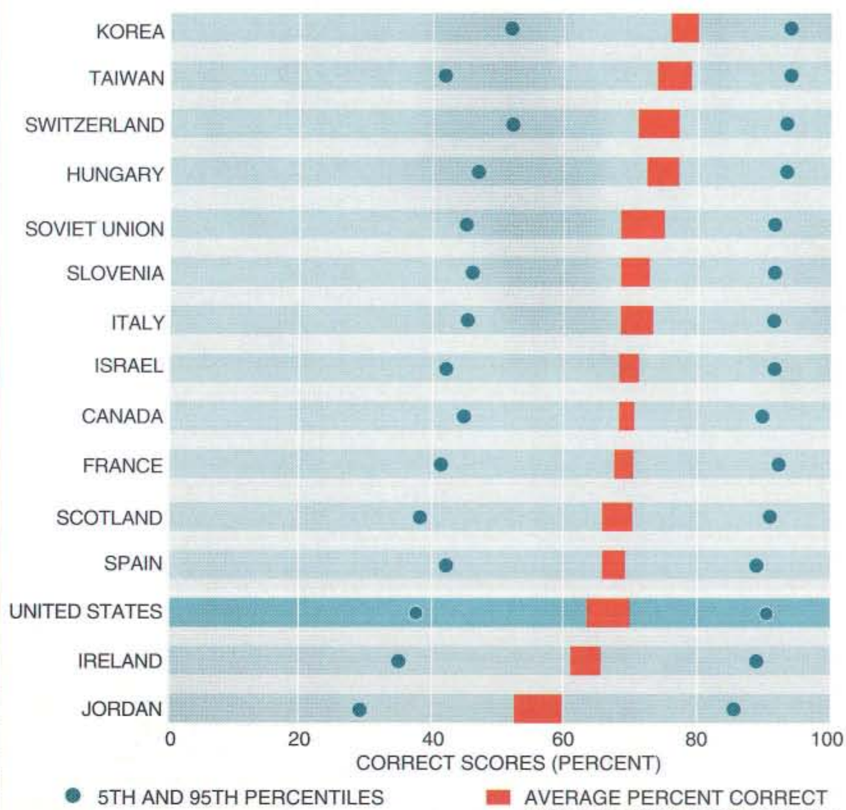
Science education programs began to flower anew in the mid-1980s, after the publication in 1983 of an influential report by the National Commission on Excellence in Education entitled "A Nation at Risk." The report prompted many states to increase high school graduation requirements in math and science, but by 1990 only four states required the three years of basic science recommended by the report. "We seem to have a penchant for having good ideas and then forgetting them," muses Senta A. Raizen, director of the National Center for Improving Science Education in Washington, D.C.

Looming behind the reform initiatives is the question of how much society is willing to pay for education. Many school districts are suffering under severe budget constraints. "You can't change when you're in crisis," says Karen Worth, an educator at the Education Development Center, Inc., in Newton, Mass., and chair of the NRC working group that is developing standards for teaching science. The median expenditure in 1989 per pupil in school districts nationwide was close to $4,000, but dozens of districts spent less than half that amount, according to the Department of Education's National Center for Educational Statistics.

In parts of Alabama and Mississippi, books and desks are in short supply, let alone computers. Poverty at home is another daunting problem. The Head Start program, which provides early learning experiences for youngsters from impoverished backgrounds, reaches less than 30 percent of at-risk children, notes Leon M. Lederman, a Nobel laureate physicist who has established an academy for inner-city science teachers who work in Chicago. The Carnegie commission estimates that by the year 2000, one quarter of children in the U.S. will be living in poverty.

A Failing Grade

U.S. students compare poorly with those in other countries in their knowlege of science. When the International Assessment of Educational Progress studied 13-year-olds in 15 countries during 1990–1991, U.S. students ranked 13th.



Textbook Watchdog

Publishers quail when they pick up *The Textbook Letter,* the bimonthly bulletin of the California Textbook League (P.O. Box 51, Sausalito, CA 94966). William J. Bennetta, the bulletin's combative editor, writes that with the exception of some books for advanced placement courses, most textbooks are "wretched junk" written by "clowns who know nothing about the subject matter." Here are a few extracts from the *The Textbook Letter*'s reviews:

Addison-Wesley Life Science 1989: It has many blatant, basic errors... when one seeks *evolution* in the glossary of Addison Wesley's book, one finds only: **"See *theory of evolution by natural selection.*"** Doing that, one sees:... "The idea proposed by Charles Darwin that organisms have changed gradually over a long period of time. Darwin believed that helpful variations developed by chance, but that individuals with these variations were then more likely to survive to bear young and thus pass on the variation." That is false.... The idea that organisms have changed over time—that they have undergone evolution—did not originate with Darwin, is not unique to Darwin, and is not the same as the idea of natural selection.... (From a review by Ellen C. Weaver.)

Physical Science: The Challenge of Discovery (D.C. Heath and Company, 1991): A book so poorly done that it should be withdrawn.... After telling us (on page 590) that "ancient people found that charcoal burned hotter than wood," Heath's writers announce: "This form of coal was used for heating in areas where wood became scarce."... *Physical Science* is a travesty of science... Page 156 says that copper is useful in making saucepan bottoms because its heat capacity is low—its "temperature rises measurably [*sic*] with only a small amount of added heat." Just four pages later, however, the reason for using copper is that its thermal conductivity is high. It seems that Heath's writers are truly befuddled.... (From a review by Lawrence S. Lerner.)

Life Science 1989 (Macmillan Publishing Company): Page 9 presents... a purely imaginary story about vitamin C. "In the 1800's," the writers say, some "biologists" conducted some "tests" and found that sailors often suffered from scurvy because "vitamin C was missing from the sailors' diet." But then, "When limes, lemons, and oranges (fruits with vitamin C) were added to their diet, the symptoms disappeared." That is all false.... The dietary prevention of scurvy...was established not by biologists but by Dutch and British naval officers; these included the famous Capt. James Cook, who lived in the 1700s (not the 1800s). Further, the writers' use of the term *vitamin C* is anachronistic and misleading. (From a review by Leighton Taylor.)

The standard-bearers of the new reform movement insist that this time things will be different. "These are going to be everyone's standards," Hoffman declares. Previous attempts to overhaul science education, he says, failed to concentrate on "systemic changes" that influence the whole educational system. "To change the system, you have to affect all of it," adds Jerome Pine, a biophysicist at the California Institute of Technology who has engineered a successful science-teaching collaboration with the Pasadena Unified School District.

The approach advocated by Pine and his co-revolutionaries emphasizes continuing in-service training for teachers rather than one-shot courses that leave them floundering when they return to their often unsupportive schools. It also places a high priority on improving the way student learning is assessed. Teachers say they deplore the practice of "teaching to the test"—drilling into students just the facts that are likely to come up in an examination. But that strategy seems to be common. "The grip of standardized testing has trivialized what counts as knowledge and denied teachers' own intelligence and good judgment," says an angry Eleanor Duckworth, a childhood learning specialist at Harvard University.

One organization called the New Standards Project is already drafting questions for a nationwide examination system and is working on criteria for calibrating other exams. The group wants to avoid relying on the multiple-choice tests that are now standard. Instead it seeks exams that assess cumulative accomplishments and performance on set tasks, as scout merit badges do.

The independence of states and local school boards is still a critical issue. They provide 94 percent of the budgets for schools and have traditionally guarded with jealousy their right to decide what is taught, as well as how it is assessed. "The concept of national standards was until recently something people didn't want to consider seriously," says Henry W. Heikkinen, a chemistry professor at the University of Northern Colorado who is chairman of the National Research Council's working group on curriculum standards. But Hoffman says, "Now I see far greater recognition on the part of the major players that they have to be part of the same effort."

The political Rubicon was crossed at an "education summit" convened by President George Bush in Charlottesville, Va., in 1989. (The gathering was chaired by Governor Bill Clinton of Arkansas.) Shamed by a dismal procession of studies that showed in painful detail how the U.S. has lagged most other countries in math and science over the past decade, the governors of all 50 states agreed to sign on to the president's ambitious list of "National Educational Goals."

Audacious would be nearer the mark: one goal, for example, ringingly declares, "By the year 2000, U.S. students will be first in the world in science and mathematics achievement." As Frank B. Murray, dean of the College of Education at the University of Delaware, points out, it is not even clear what the goal of being first means; simply figuring out how to assess learning is controversial.

Most teachers would be pleased with any sustained improvement, and educators have taken a generally pragmatic view of the goals. Many seem to have accepted the rationale offered by Diane S. Ravitch, assistant secretary for educational research: "If you work toward goals, you'll be a lot closer to them," she says. "The American system has never had goals."

Teachers, scientists and politicians are, however, far less sanguine about the strategy the "Education President" offered last year for meeting the goals, under the banner of "America 2000." Indeed, many educators see the program as the result of cynical political calculations rather than a serious anal-

ysis of needs. Of the $770 million that the administration requested for America 2000 in 1993, for example, $500 million is tagged to help states provide vouchers that parents could use toward the cost of a private school. "It's a joke," says Bill G. Aldridge, executive director of the National Science Teachers Association and the driving force behind its effort to reform the science curriculum. "Why didn't they put that money into poor urban school districts?"

Another high-profile component of America 2000 is a plan to establish at least 535 "model schools," one or more in each congressional district. Rollin Johnson, a consultant to the Carnegie commission, points out that districts vary enormously, so putting a new school in each one cannot be an efficient way to effect change. A mere 250 of the 16,000 school districts in the country educate 30 percent of the work force. In any case, "just because you've built a few hundred model schools doesn't mean you've dealt with the problem," Raizen says.

Yet along with the voucher scheme and the model schools, America 2000 also called for national standards in key disciplines, including math and science. Whatever the misgivings about the overall thrust, teachers and academics realized the standards presented an opportunity. They decided they had better get on board or get left behind. Frank Press, president of the National Academy of Sciences, secured the contract for coordinating development of the standards for the NRC, the academy's operating arm.

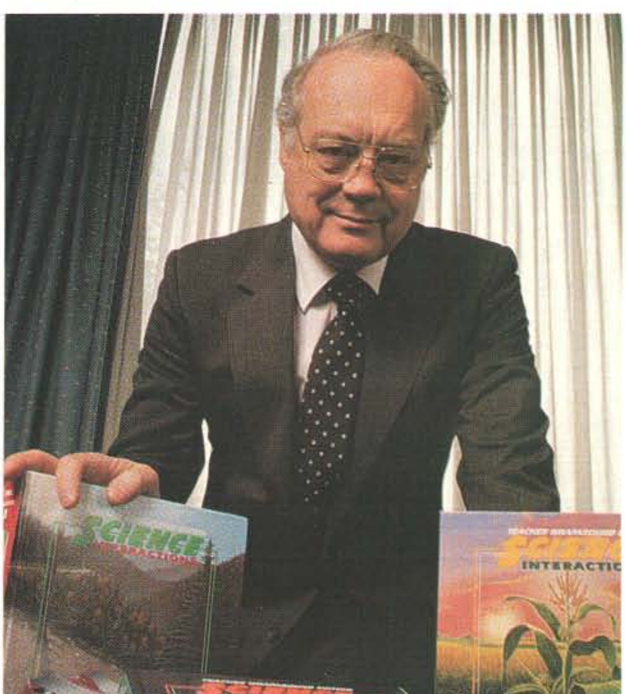
The science curriculum standards that Heikkinen's working group is now fleshing out will leave local jurisdictions plenty of flexibility, says presidential

Impatient Reformer

Bill G. Aldridge, executive director of the National Science Teachers Association, is a determined ex-physicist who was the force behind the association's curriculum reform campaign, Scope, Sequence and Coordination. Aldridge came close to lining up his association to design national curriculum standards, before bowing out to let the National Research Council take charge. But he has been criticized for writing spin-off textbooks from Scope, Sequence and Coordination before the project is completed.

Aldridge's uncompromising stand on teaching basic science has also led him into conflict with supporters of a movement known as Science, Technology and Society (STS). This philosophy seeks to make science palatable by presenting it with societal issues. "We put the student at the center of the problem," explains Robert E. Yager, an STS advocate at the University of Iowa who is adapting Aldridge's core curriculum. But Aldridge has no time for diluted science. "When I was teaching density, one third of middle school teachers could not use the gradations on a ruler to measure fractions of an inch," he complains. "They were the same ones who wanted to go off and measure acid rain."

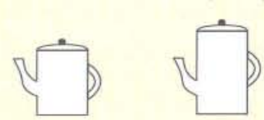
Aldridge argues that traditional test questions fail to probe real understanding. A typical example might be: Which of the following gases must an animal breathe in order to remain alive? A. Helium. B. Hydrogen. C. Nitrogen. D. Oxygen. Aldridge has devised some very different questions [*see below*]. But Rustum Roy, a professor of materials science at Pennsylvania State University and a supporter of STS, retorts that Aldridge's questions are suitable only for future Ph.D.'s. "He is trying to make better scientists, but we don't need any more scientists," Roy fumes.



BILL G. ALDRIDGE emphasizes understanding.

1. It is 11 P.M. in late September. When you look west, slightly southwest, you see the moon just above the horizon (about 15 degrees). Sketch a picture of how the moon appears. If it is not a full moon, show which part of the visible disk is dark and which part is illuminated by the sun. Where will the moon be the next night at the same time? Higher or lower in the sky?

2. The illustration below shows two coffeepots sitting on a level tabletop on the earth. Both are cylindrical and have the same cross-sectional area. Which coffeepot would hold the most coffee? Explain your answer.



3. Suppose that you weigh a glass full of water. Then you weigh an identical glass that is full of water but has a wooden block floating in it. Which weighs more? Explain your answer.

[*Answers on page 86*]

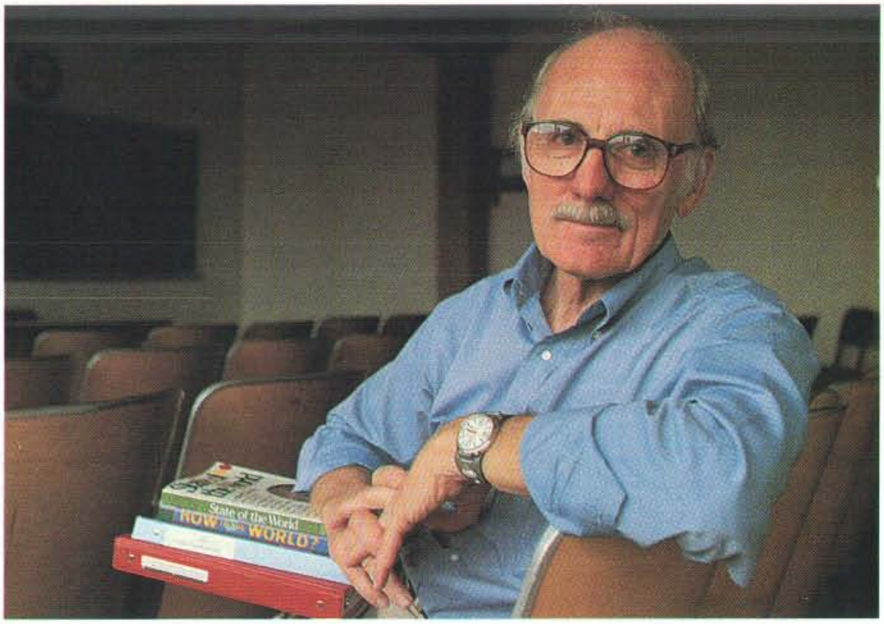
science adviser D. Allan Bromley. They will not, for example, specify what students should know down to each grade level. Rather they will offer a consensus on what students in broad age brackets ought to understand. To try to coerce states into following a rigid national curriculum would, Bromley says, simply slow the adoption of the standards.

The federal government will nonetheless be able to apply some not so subtle pressure to states to keep them in line. Recipients of competitive grants for science education projects awarded by the Department of Education, Ravitch says, will have to demonstrate that their projects conform to national criteria. States are now using some of those grants to develop their own curriculum standards.

Nobody believes national standards will bring the goal of being the world leader in math and science by 2000 within range, any more than the goal that says every adult American will be literate by the year 2000 or the one that says every school will by then be free of drugs. Time is simply too short: the high school graduates of the year 2000 are already in fifth grade. And presumably they are already falling behind their contemporaries in Korea, Taiwan and Switzerland, who took the top three places in a recent comparison of math and science ability in 13-year-olds conducted by the Educational Testing Service. Entrants from the U.S. ranked 13th out of 15 countries sampled; only youngsters from Jordan and Ireland fared worse. "Not everyone is happy about the way these comparisons are done, but it's embarrassing to come up short again and again," says Lapp of the National Science Resource Center.

International comparisons are not necessary to see U.S. high schools flunk in science education. A survey of 12,000 high school students in 1988 found that half of those who had never taken a course in biology did as well in tests as 40 percent of those who had; apparently, biology courses taught most of those taking them almost nothing. Moreover, "The 1990 Science Report Card" concluded that less than one half of high school seniors could apply scientific knowledge to interpret data, evaluate and design experiments, or display some in-depth knowledge of science.

Although achievement scores improved a little during the 1980s after declining during the 1970s, performance in 1990 was no better than it was in 1970. Luther S. Williams, who heads Education and Human Resources at the National Science Foundation, estimates that employers in the U.S. now spend in the region of $100 billion a year retraining high school graduates in basic skills.



F. JAMES RUTHERFORD directs the American Association for the Advancement of Science's Project 2061, a long-range plan to rethink how science should be taught.

Still, national standards should eliminate some of the wide variation among states, if the experience with math standards is any guide. The National Council of Teachers of Mathematics published universally applauded national standards last year, and many states have already drawn on them for their own curriculums.

The Attitude Problem

Whether more consistent state and local requirements can improve education in the absence of more far-reaching social changes is an open question. Those who have studied learning in different cultures are convinced that societal attitudes are partly to blame for U.S. students' distaste for math and science. The subjects have the reputation of being difficult. And surveys by Harold W. Stevenson and his colleagues at the University of Michigan suggest that in both Japan and China, students and their parents attach greater importance to academic achievement than do their counterparts in the U.S.

According to John H. Bishop, who studies education at Cornell University, the root cause of U.S. students' poor performance is the "absence of good signals of effort and learning in high school." In the U.S., unlike other countries, the standardized tests taken by seniors measure aptitude rather than how much has been learned. The primary indicators of learning achievement, in contrast, are grades and class rank, which are only relative measures.

Bishop suggests that ability to pay, rather than achievement, is what mainly determines access to higher education in the U.S. Therefore, he argues, students have little to gain from working hard, and parents have little incentive to vote tax increases for schools. "No adolescent wants to be considered a nerd, brain geek, grade grubber or brownnoser, yet that is what happens to students who study hard and are seen to study hard," he concludes.

One consequence is that few high school students continue in science. Biology is the only science course that most high school students ever take. Less than 50 percent of the graduates of the class of 1990 took chemistry, and only about 20 percent took physics. In 1985, 19 percent of high schools did not even offer physics, and 9 percent did not offer chemistry.

As long ago as 1916, John Dewey was emphasizing the importance of giving all students—not just a select few—a basic scientific education. Yet most children, surveys show, are "turned off" to math and science at an early age. Usually they are bored or intimidated by lessons that consist mainly of rote memorization of words that are useful only for passing tests. Girls and members of minority groups—with the exception of Asians—are especially likely to reject math and science.

Attitudes are hard to change, but there is little disagreement about how math and science ought to be taught, at least in the lower grades. Educators since Dewey have been stressing the importance of allowing youngsters to do experiments and make observations, in

order to give them a taste of the excitement that comes with discovery. And a strong consensus holds that interest in science is most easily sparked in the elementary or middle school years.

Yet at that level, many school systems assign science a low priority. Most elementary teachers have no formal training in the subject. As if that were not bad enough, many elementary schools make it hard for teachers to get the materials they need for hands-on science. The situation is not much better in many middle schools. And high schools often have to rely on out-of-date teaching materials, Lapp says.

In addition, high school students are faced with an unappetizing layer cake: a year of biology followed by a year of chemistry followed by a year of physics. Few educators are prepared to defend that arrangement, which is peculiar to the U.S. "The sequential system used in the U.S. deprives students of the opportunity for integrated learning," the Carnegie commission reported last year. And performing an experiment is a rare event for many students even in high school: 26 percent of grade 12 students who are taking science say they never do experiments, according to "The 1990 Science Report Card."

The reasons are not hard to discern. Assembling materials and equipment for practical demonstrations takes time and effort and can be expensive. Teachers find it easier to rely on textbooks for their lessons instead. Although textbook publishers say they are attempting to incorporate some experimental flavor, an NRC report on biology education noted two years ago that "most professional biologists who examine high school texts are appalled at what they find." The books "are often boring, and they are also sometimes either misleading or incorrect." Small wonder that so few students elect to take another science course after biology. They have a point—physics and chemistry textbooks are, by all accounts, just as bad.

To many educators, the solution lies in a philosophy of learning called constructivism that was espoused in the 1983 "A Nation at Risk" report. The idea draws on the work of child psychologist Jean Piaget, but it can be traced back to Dewey and the turn-of-the-century school called progressivism, which emphasized building bridges between subjects. Constructivism holds that children must build their own meanings in order to understand, and its mantra is "less is more." Understanding, in this view, is best helped by cutting extraneous facts out of the curriculum and eliminating rote memorization. Instead students are encouraged to explore "rich multiple representations of a concept," says Murray of the University of Delaware.

The challenge is in putting that idea into practice. "Teachers can have messy, hands-on classrooms, but it doesn't necessarily mean any science gets taught," Raizen observes. "Teachers have to understand the basic science involved." Hands-on elementary science courses developed 30 years ago were big on exploration but demanded a good deal of knowledge on the part of teachers. "I think in the 1960s we were psychologically and pedagogically naive," says Robert Tinker of Technical Education Research Centers in Cambridge, Mass.

Several efforts to avoid the pitfalls and lay new frameworks for the future are now getting under way, in addition to the standards projects at the National Research Council. "It's, 'let's all get into a taxicab and open our umbrellas at the same time,'" Raizen jests. One of the most energetic is Scope, Sequence and Coordination, the program started by Aldridge of the National Science Teachers Association. Six test areas, including 34 middle schools in Houston and several hundred other schools from Alaska to Puerto Rico, are now developing and testing teaching materials based on that project's "core" guidebook. The document rejects the traditional biology-chemistry-physics approach. Instead it sketches time lines for teaching all of the sciences concurrently in middle school and high school.

The grandest of the reform efforts, and the great rival of Scope, Sequence and Coordination, is a plan by the American Association for the Advancement of Science known as Project 2061. F. James Rutherford, its director, probably rues the day the project received its cryptic name, which refers to the year when Comet Halley will next return to the inner solar system. The moniker has made the 25-year effort the butt of jokes along the lines of 2061 being the year when the project will begin to influence teaching. "I don't see Project 2061 making any significant change in time to be useful," gibes Aldridge, who before he handed over the reins two years ago led Scope, Sequence and Coordination at what some saw as a breakneck pace.

Rutherford patiently cites the fable of the tortoise and the hare and explains that the thorough reexamination of science teaching being undertaken in Project 2061 is not something to be hurried. "Grass-roots reform didn't work, and top-down reform didn't work. You can't just do a curriculum and insert it," Rutherford says. Nevertheless, he insists: "We'll get there first."

Project 2061, in fact, won't produce a curriculum, in the sense of materials that a teacher can use directly to plan a course. That task will be left to others. But six Project 2061 teams of 25 teachers each are now working hard at thinking up entirely new approaches to science education. Questions such as alternatives to textbooks and different ways to structure the school day are being explored and described in "curriculum models." "We're connecting to the arts and humanities, building a new scientific culture," Rutherford says with feeling.

Project 2061 has already published summaries of what it has decided high school graduates should know about different scientific subjects. These ideas, according to Rutherford, have been influential in helping districts and states craft their own curriculums and have attracted interest abroad. And in a politically savvy move, Project 2061 recently turned to producing "benchmarks" that suggest what should be known by different age groups. That work has been folded into the NRC's curriculum standards effort and into California's curriculum "framework."

STUDENTS erect an osprey platform at St. Andrew's, a private school in Middletown, Del., that has its own lake (*left*). Public school students rarely have such

Project 2061's studies have convinced Rutherford that conventional textbooks are becoming too big and expensive and "have had their day." One widely hailed solution to the textbook problem is the computer. The National Geographic Kids Network, for example, which was developed in conjunction with Tinker of Technical Education Research Centers, is highly regarded. It allows students in different locations to share and compare experimental data. But Jan Hawkins of Bank Street College in New York City, who has written acclaimed educational multimedia programs, says most commercially available educational software is far less imaginative than the Kids Network. Hawkins notes that many programs fail to make students think deeply and are used as substitutes for hands-on science. At the Computerworld Smithsonian Awards this past summer, some of the educational entries were little more than high-tech versions of flash cards.

Seymour A. Papert, a computer education expert at the Massachusetts Institute of Technology's Media Lab, believes the use of computers is deteriorating as computers are increasingly segregated in special rooms. And he warns that because so many school districts are now cutting back on expenditures, opportunities to make good use of computers may become even scarcer.

opportunities. Even if they have computers, the machines typically run unimaginative software and are segregated from classroom activity (*right*).

Meanwhile the search for educational innovations is moving ahead, albeit fitfully. A not-for-profit company called the New American Schools Development Corporation, a spin-off of America 2000, has funded 11 private consortia to design and build "break the mold" schools that it hopes will provide a better education at the same cost as today's institutions.

But the corporation has published scant information about the concepts it has funded even though the first schools are supposed to open next year and provide ideas for America 2000's 535 model schools. And although President Bush challenged the corporation last year to find $150 to $200 million for the work, it has so far raised only $50 million.

Voucher Game

Nevertheless, some hope the private sector will deliver. If the administration's voucher scheme were to be enacted by Congress—which staffers say is unlikely in its present form—the fortunes of private schools could be boosted. H. Christopher Whittle, the mercurial president of Whittle Communications in Knoxville, Tenn., may be riding the wave.

Whittle, who is best known for his commercially sponsored "Channel One" classroom television news programs, has announced plans to set up a network of private schools that would use fresh ideas to provide educational excellence at no more cost than do public schools (as well as give away scholarships). But Benno C. Schmidt, Jr., the former president of Yale University who joined Whittle earlier this year, is unable to explain how the proposed schools will pull off that feat, and many educators are skeptical. Nancy Hechinger, the science education specialist at Whittle, says she is thinking of using parents as a teaching resource but has little more to offer by way of specifics.

Any effort to improve science education—or education in general—will be moot without skilled teachers. Franklin M. Loew, dean of the veterinary school at Tufts University, and John B. Slaughter, a former director of the National Science Foundation who is now president of Occidental College in Los Angeles, both believe increasing the number and the quality of high school science teachers is the most urgent priority.

One place to start would be the science courses that most high school teachers take as part of a liberal arts degree, Loew argues. Many of those courses are failing to provide tomorrow's teachers with understanding, he says. More than two thirds of math and science teachers fall short of standards for course-work preparation set by professional associations, according to the Carnegie commission.

Alan B. Krueger, an economist at Princeton University, suggests that reducing class sizes and increasing teachers' pay would be a better use of federal funds than America 2000. Until recently, advocates of better pay as a way of improving the teaching force were embarrassingly short of data to support their case. But Krueger points to a new study by Ronald F. Ferguson, a public policy scholar at Harvard University, which supports that strategy. Ferguson examined almost 900 school districts in Texas and found that teachers who performed well on the statewide recertification exam were likely to be employed by districts that paid higher salaries. Not surprisingly, their students scored higher on standardized tests, even after allowance was made for socioeconomic differences.

Meanwhile nationwide initiatives aimed at improving science education are recognizing the importance of continuing training for teachers. Many are collaborations involving professional scientists at universities or federal laboratories. Bromley notes that education is now officially a part of the mission of the government's science-based departments. The Department of Energy, for example, has since 1990 been following the lead of the National Aeronautics and Space Administration by encouraging its laboratories to forge partnerships with local schools. Teachers and students visit to see exhibits and participate in research, while laboratory personnel are farmed out to advise in schools.

At the Teachers Academy for Mathematics and Science in Chicago, founded in 1990 by Lederman, teachers from inner-city schools come for retraining two days out of every two weeks. On a third day they are helped to use their new knowledge by a mentor teacher back in their own classrooms. Follow-up is considered an essential part of the program.

Other projects use technology to provide support for teachers. Tune In Math and Science, a teaching enhancement program run by the GMI Engineering and Management Institute in Flint, Mich., uses audio and video links to allow remote classes to hook up with

LEADERSHIP INSTITUTES run by the National Science Resource Center in Washington, D.C., help teachers and others become familiar with hands-on experiments. Participants this past July practiced building electric motors.

mentor teachers. In the collaboration engineered by Pine at Caltech, known as Project SEED, scientists work alongside teachers with teaching materials but do not lecture. Master or mentor teachers spread the word about good practice.

In addition, efforts to put better materials into the hands of teachers are beginning to pay off. The NSF has stepped up its efforts to develop up-to-date curriculum materials for the higher grades. And in the elementary grades, Lapp says, modular science kits developed at the National Science Resource Center and elsewhere provide more guidance for teachers and are more easily refurbished than are the kits of 20 years ago. The "leadership institutes" held at the center give teachers "a little sphere of confidence" using the kits, Lapp says. School districts wanting to send teachers to the institutes must send administrators and scientists as well, to build support for hands-on science.

Bromley points out that federal outlays on precollege math and science education programs have more than doubled in the past three years. With the rubric of a "presidential initiative" to smooth the way, the administration has proposed spending $768 million in 1993, in addition to America 2000. Expanded, intensive in-service training programs are planned to reach almost half of all math and science teachers in the country. The NSF has inaugurated a special Statewide Systemic Initiative to encourage improvements in science education. Grants available through the initiative are meager, however. They amount to only $10 million per state over a period of five years. "That's about the paper-clip budget in California," Pine comments.

Congress has in the recent past been generous to science education, although rigid budget constraints put this year's proposed allocations in doubt. Bromley says that the U.S. spends more than any other country in the world on education (possibly excepting some Swiss cantons) and that an attack on bureaucratic waste will free enough resources to bring science education into line with the new national standards.

But many others are now convinced that effective reform cannot be done on the cheap. The proposed federal expenditures in precollege math and science education still account for less than 8 percent of what the Department of Education plans to spend on all precollege education next year. Over the long term, says Hoffman of the National Research Council, "some significant increase in resources is going to be necessary for the success of most of the changes people are talking about."

Answers to questions on page 82

1. If the moon were in the southwestern sky just above the horizon at 11 P.M., it must have been approximately overhead (toward the south) at sunset, five hours earlier. So the sun was, and still is at 11 P.M., shining on half of what appears to be a moon disk. The illuminated half of the disk is to the right. Since the moon revolves around the earth in the same direction as the earth's rotation, on the next night the moon should be higher in the sky.

2. The spouts are at the same level, so the liquids in the two pots will move under pressure to the same level because any excess will spill from the spouts. Since the pots have the same cross-sectional area, they must hold the same amount of coffee.

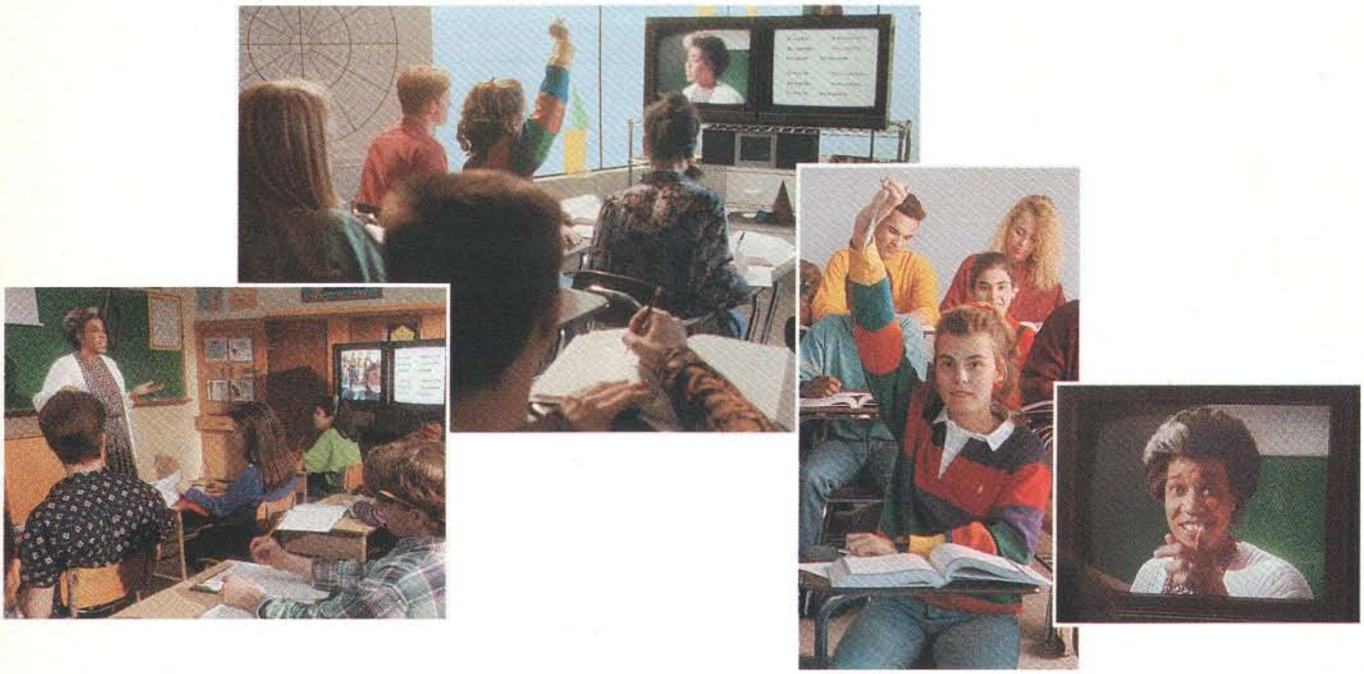
3. The block displaces an amount of water equal to the block's weight. This water overflows the glass, so the weight of the two glasses of water is the same.

FURTHER READING

IN THE NATIONAL INTEREST: THE FEDERAL GOVERNMENT IN THE REFORM OF K-12 MATH AND SCIENCE EDUCATION. Task Force on K-12 Mathematics and Science Education. Carnegie Commission on Science, Technology and Government, September 1991.

THE REFORM OF SCIENCE EDUCATION IN THE U.S.A.: DÉJÀ VU OR DE NOVO? Senta A. Raizen in *Studies in Science Education*, Vol. 19, pages 1–41; 1991.

STRATEGY FOR ACHIEVING EXCELLENCE IN SECONDARY EDUCATION: THE ROLE OF STATE GOVERNMENT. John H. Bishop. Working Paper No. 91-24. Center for Advanced Human Resource Studies, Cornell University, 1991.



Storage Space

Imaging technologies offer promising ways to store data

Like city dwellers trying to stuff a lifetime's belongings into a tiny closet, computer researchers are wrestling with how to store ever more data in ever less space. This summer, researchers at AT&T Bell Laboratories and the IBM Almaden Research Center have proved as resourceful as any New Yorker as they pack dramatically more data onto the same disk space.

In July, a Bell Labs team in Murray Hill, N.J., described how it had fit about 100-fold more data—45 gigabits (or billion bits) per square inch—than are typically crammed onto a magneto-optic disk. (That amount of data is equivalent to storing two copies of *War and Peace* in an area the size of a pinhead, AT&T asserts.) In late August, IBM researchers in San Jose, Calif., were scheduled to publish a paper describing how they can pack 30 gigabits per square inch of data onto a write-once plastic disk. The workers claim their system can read those data about 10 times faster than can the system built by their East Coast competitors.

The work is attracting attention both for innovation and for its commercial promise. Data storage already amounts to a worldwide $40-billion business. The growth of high-definition television, ever more sophisticated telecommunications systems and computer networks will boost the demand for data storage. As a result, competition among researchers—even at institutions usually noted for their dignified demeanor—is growing fierce.

To inscribe so many data in such a small area, Bell Labs, IBM and others have begun exploiting tools more frequently used to portray minuscule structures than to create them. Bell Labs scientists, for instance, are relying on a descendant of the classical optical microscope, whereas IBM's team has turned to a more recent technology, the atomic force microscope (AFM). Both methods use a so-called scanning probe, a device popularized by the success of the scanning tunneling microscope (STM), which won a Nobel Prize for two IBM researchers in 1986.

An STM typically relies on a tiny probe suspended a constant distance above the surface of a sample to detect surface contours and bumps as small as a single atom. Researchers then measure variations in a tiny electric current between the probe and the sample. Although STMs still hold the record for detecting the smallest structures, they operate too slowly to be used as a tool for recording data.

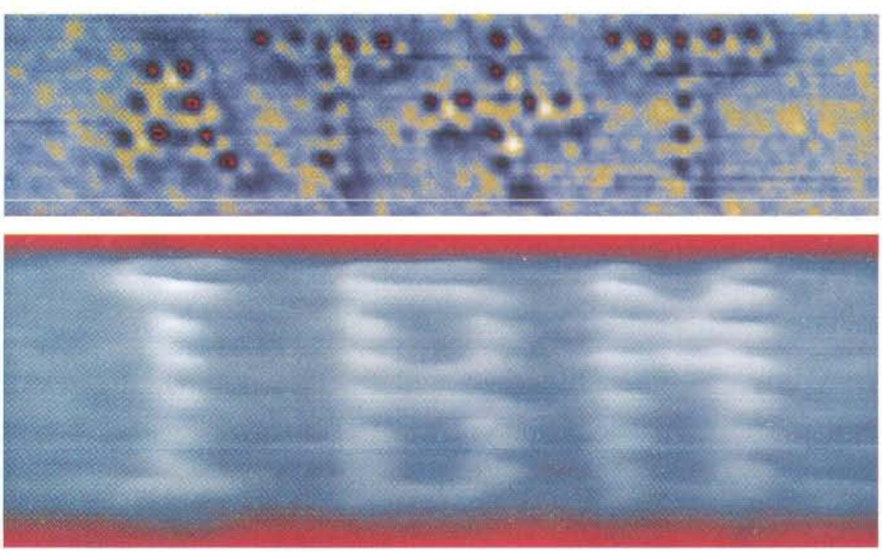
STMs nonetheless inspired the work of scientists trying to sharpen the abilities of a technology already centuries old, the optical microscope. The wavelength of visible light constrains the resolving power of conventional optical devices. Yet decades ago scientists realized they could extend those limits if they could put a small light source a few fractions of a wavelength from the sample. More recently, researchers added an STM-like probe that could be moved across the surface of the sample, thus opening up a new discipline, near-field scanning optical microscopy.

To achieve the record-breaking storage densities, Eric Betzig and Bell Labs colleagues Jay K. Trautman, Ray Wolfe, E. Mike Gyorgy and Patrick L. Finn wove together three advances. First, they improved their probe: by stretching a fine optical fiber until it broke, then plating the glass with a metal sheath, they created a 20-nanometer-wide probe that efficiently channeled light onto a small spot.

Next the Bell Labs scientists developed a technique for maintaining a constant distance between the probe and the sample, called sheer force microscopy. Like a bristle on a brush, the probe is held perpendicular to the surface of the sample. As the probe is vibrated back and forth, the researchers measure the drag caused by air between the probe and the sample. In contrast, the probe of IBM's atomic force microscope directly touches the surface of a disk, as a stylus of a record player does. (It makes pits on the surface when heated by a laser.)

Finally, the group made use of a magneto-optic film developed at Carnegie Mellon University for high-density storage. The film was not designed for Bell Labs' work, Betzig points out. "We got excited about our approach because it really borrows heavily from existing media," he says. "This gives us a head start on the road to commercialization."

To read an area of a disk, the Bell Labs researchers mount the fiber probe on a scanning device resembling those found on conventional disk drives. By pumping in low-power laser light, they can read the pattern of dots on the surface, each of which measures between 60 and 80 nanometers in diameter. A more high-powered laser "writes" data



***HUNDREDFOLD MORE BITS PER INCH** can be written multiple times on a magneto-optic disk by researchers at AT&T, using a scanning optical microscope* (top). *With an atomic force microscope, IBM has achieved a similar storage density on a write-once plastic disk* (bottom). *Many engineering challenges must be met before such tools reach the market.*

by reversing the magnetization of a spot. The change in polarity of the spots alters the way they reflect light. The physics behind the writing phenomenon is still unclear, Betzig observes. His favorite explanation: the laser heats the tip of the probe, which reverses the magnetism through thermal effects. The disk can consequently be rewritten many times.

Atomic force microscopes, on the other hand, can write and read data faster than can their optical cousins, IBM researchers contend. In the August 31 issue of *Applied Physics Letters*, Daniel Rugar and H. Jonathan Mamin are scheduled to report that the heated tip of an AFM can read and record data on a plastic disk at rates up to at least 100 kilohertz. (One disadvantage is that the data are permanently recorded on the disk.) In contrast, the Bell Labs tools now read data at a rate of about 10 kilohertz, a fraction of the speed of a conventional compact disc player.

"The read speed is the real challenge," Betzig concedes, because it is limited by how many photons can squeeze out of the probe at a time and so is limited by the wavelength of laser light. Because an AFM simply measures the pits in the substrate surface, it can easily achieve higher reading speeds. Nevertheless, "we think we could get 10-megahertz speeds, for both reading and writing," Betzig asserts, "but it's proprietary."

Reading and writing are only the first steps toward machines that handle more data. Many engineering challenges await, including detecting and correcting errors and keeping track of a specific bit in the sea of data. "If everything went right," Betzig muses, "the best we might achieve would be a product about five years from now." IBM is even more taciturn. "Obviously we're interested in it because of the potential for an important technology impact," says Rugar, manager of exploratory storage studies. But he remains reluctant to speculate when IBM might try to move the research closer to the market.

There is, moreover, a spectrum of other technologies, including different varieties of scanning probes, that promise to increase data storage. At Stanford University, for instance, Calvin F. Quate and his colleagues are using combined force and capacitance microscopy to pack data onto a disk. Elsewhere, investigators continue to explore three-dimensional storage techniques, such as holographic crystals.

Betzig remains enthusiastic. "There are zillions of ways to store data," he notes cheerfully. Now engineers must demonstrate which techniques are most practical. —*Elizabeth Corcoran*

Different Strokes

A two-stroke engine without the blues

The humble two-stroke engine, long prized for its small size but denigrated for its blue smoke, has taken on a surprising new role as environmental savior. It owes this makeover largely to the Orbital Engine Corporation of Perth, Australia, a small research firm that has revised the century-old design, taking it out of the lawn mower and into the big time: the automobile. There it churns out more horsepower per pound, burns less fuel per mile and spews out less pollution than do conventional four-stroke engines.

The last specification is crucial to the engine's success. No auto manufacturer would invest in an engine that could not meet the tough new emissions standards in California (and other states that plan to follow its lead). These rules begin to bite in 1997, when 2 percent of new cars must meet them. The figure rises to 15 percent by 2003. An even more draconian standard, designed to encourage alternative vehicles such as electric cars, calls for zero emissions beginning with 2 percent of sales in 1998.

Orbital passed the first barrier early this year when the U.S. Environmental Protection Agency confirmed that a Ford test car powered by an Orbital engine met the California emissions standards at low mileage. The next step, says Peter Simons, president of Orbital's Michigan-based manufacturing subsidiary, is to repeat the performance over the full 50,000-mile EPA test required to ensure that the engine and exhaust systems hold up under use.

As recently as 1987, the big auto companies of America, Europe and Japan were saying they could beat Orbital at its own game. Although most of them continue to work on two-strokes independently, many have conceded that they will adopt Orbital's patented technology. Both General Motors Corporation and the Ford Motor Company, for example, have recently announced that they plan to use Orbital technology in some cars by the 1995 model year.

The principle of the engine dates back to an idea of Karl Benz, of Daimler-Benz fame. Whereas a four-stroke engine uses one motion of the piston to charge the combustion chamber, another to compress the fuel-air mixture, a third to develop power from the resulting explosion and a fourth to expel the combustion products, a two-stroke engine makes twice as many power strokes, producing nearly twice the ratio of power to weight. A GM test engine incorporating Orbital technology tips the scales at only 165 pounds, compared with about 250 pounds for an equivalent four-stroke engine.

The weight savings, together with secondary reductions in the supporting structures and brakes, adds up to improved fuel economy. Orbital says cars equipped with its engines consume between 5 and 35 percent less fuel than do comparable cars with four-stroke en-

GREEN MACHINE: Peter Simons, head of Orbital Engine's U.S. subsidiary, shows off the Ethos, which carries the firm's fuel-efficient engine. Photo: Peter Yates.

gines. Company representatives maintain that a car designed specifically to take advantage of the Orbital engine's compact size could save still more.

Conventional two-strokes powered Saabs and other small cars until the 1970s, then disappeared. The reason was that they polluted too much, by letting a little fuel seep into the exhaust and mixing lubricating oil directly with the fuel. Unburned hydrocarbons tinged the exhaust blue in East Germany's tiny Trabant, the last car powered by the old-style two-stroke engines. Now they are found only in lawn mowers, chain saws, outboard motors and other products in which weight is at a premium.

Orbital attacked the problem with brief, precisely timed injections of a fuel-air mixture and a separate lubricant injection system that keeps oil out of the exhaust. It then strove to keep up with tightening automobile emissions standards by refining the engine and the catalytic converter that oxidizes pollutants in the exhaust.

Orbital says it has a chance of meeting the zero-emissions standard. "We're doing work to get what we call 'near zero' emissions," Simon says. "If you take into account that energy used to generate electricity is not itself zero, we should be competitive" with electric cars.

The company plans to reduce hydrocarbon emissions to concentrations below those in the ambient air of Los Angeles and New York. "It's an intriguing concept," Simon observes. "You'd have cars driving around, cleaning the air." *—Philip E. Ross*

Orphan Isotopes

Scarce isotopes bring meager profits to the DOE

Many physics and chemistry experiments share a need with technologies as diverse as smoke detectors and some medical imaging approaches: they rely on stable or radioactive isotopes produced in separators called calutrons or in nuclear reactors and particle accelerators. These days, isotopes are also at the heart of an experiment in economics—but one that has yet to produce a successful result.

According to the free-enterprise doctrines of the Reagan-Bush era, the market offers a handy solution to almost any economic issue. So why not make the Department of Energy's isotope program pay its own way? After receiving a nest egg of $16 million two years ago, the isotope program was directed to meet its future costs with its revenues. The DOE offers a cornucopia of several hundred varieties of isotopes. But because its clients are mostly basic researchers who buy or borrow only small quantities of materials, the program's revenues are meager—only $12 million in 1991. Hence, the effort is teetering on the edge of bankruptcy and hoping for a congressional reprieve.

The worldwide business in isotopes, in contrast, looks robust. Private industry manufacturers make a more limited range of isotopes, focusing on the ones widely used in nuclear medicine and commercial enterprises. The revenues for these top $500 million annually. And therein lies the rub. If the DOE drops out of isotope production, researchers worry, the types of isotopes available will fall. "If there is money to be made, then private industry can handle it," says Steven W. Yates, a chemistry professor at the University of Kentucky. "But a lot of isotopes aren't profitable. So now they might not be available for research—or will cost millions of dollars."

At a workshop held earlier this year at the National Research Council, scientific and medical researchers who use the isotopes complained about the growing scarcity of supplies and the ballooning prices. "In every field, people had problems," declares Richard L. Hahn, a nuclear chemist at Brookhaven National Laboratory who co-chaired the meeting.

Medical researchers say they have already been jarred by the vagaries of commercial production. A threatened strike at Nordion International in Ontario, the only North American-based supplier of medically important molybdenum 99, had concerned U.S. doctors looking for potential alternative suppliers as far away as Russia and Indonesia. Because the isotope is so short-lived, a strike that lasted a week would have left U.S. radiologists without the materials needed to diagnose a range of illnesses, including cancer, AIDS and heart disease.

Physicists argue that scientists face even greater difficulties. "I'm not worried about commercial users; they have the money to pay for what they need," Yates says. "But basic research is going to suffer." Since the DOE began trying to boost its revenues two years ago, Yates has found himself paying eight to 10 times more for the same materials. And even such price increases might not be enough. "It's not clear that these fees will pay the cost of the program," he says. Moreover, reserves of some isotopes are being depleted. "Each year five to 10 of these isotopes are just going out of stock," Yates adds, as the DOE cannot afford to replenish its supplies. And even as the government is putting the finishing touches on two new physics research facilities, it may lack the isotopes needed to make full use of them, says Lee L. Riedinger, a physics professor and vice-chancellor for research at the University of Tennessee.

Several facts of federal life have complicated the DOE's efforts to run a tidy business. Because government agencies are barred from competing with commercial producers, the DOE does not produce isotopes that might be profitable. In addition, it has a tough time controlling the costs of making isotopes, a task handled at the national laboratories. Programmed into the newer prices of these materials, for instance, is a prescribed share of laboratory overhead. And since the national labs have many other responsibilities, isotope production is often delayed by the priorities and glitches of unrelated programs.

Then there is the post-cold war conundrum. Like the U.S., the former Soviet Union had built up an impressive infrastructure that produced a smorgasbord of isotopes. In theory, the U.S. government should be pleased to see some friendly competition as the Russians enter the market. In practice, however, the U.S. government and industry are complaining about unfair competition. Donald E. Erb, the DOE's director of isotope production and distribution, points out that his revenues have dropped by about 25 percent because Russian entrepreneurs have been willing to sell isotopes at almost any price.

Solving the shortfall of isotopes and dollars will demand some creative management. Representative Michael L. Synar of Oklahoma held hearings on the general problems of the isotope program in midsummer. Erb hopes to find additional funding to refit a reactor at Los Alamos National Laboratory to produce molybdenum 99. If he is successful, the reactor may begin producing material by 1993, he says. Meanwhile the nuclear medicine community is also petitioning Congress for funds to study the feasibility of building a National Biomedical Tracer Facility, which would churn out a spectrum of radioisotopes for medical use and serve as a training ground for radiochemists.

Erb has a very personal reason for wanting to see the isotope program back on its feet. Seven years ago a stress test relying on the radioisotope thallium 201 showed him to have clogged arteries. "I may be alive today," he muses, "as a consequence of that test." *—Elizabeth Corcoran*

Fine Grain

Nanophase materials move out of the lab

A few years ago Larry E. McCandlish, then a researcher at the Exxon Research and Engineering Company in Annandale, N.J., began working with his colleague Richard S. Polizzotti on finely powdered metal carbides. McCandlish was initially interested in the materials as potential catalysts. But he soon realized that these powders could have useful physical as well as chemical properties.

Like other materials engineers, McCandlish was intrigued by the possibility of "nanophase" materials. These are ceramics, metals and combinations of the two that are made from extremely small particles, less than 100 nanometers in diameter. They should have unusual properties because so many of their atoms lie on the edges of the tiny "grains." So McCandlish moved to Rutgers University, where, together with Bernard H. Kear, head of the nanophase materials research laboratory there, he started to search for new ways of making ultrafine powders of cobalt tungsten carbide, a material used in cutting tools and other wear-resistant devices.

BERNARD H. KEAR with glass spray-dryer that he uses to make "nanophase" mixtures of metal salts at Rutgers University. The operation is his first step toward creating materials of exceptional hardness. Photo: Robert Prochnow.

Two results of their investigation are a process for making nanophase particles of cobalt tungsten carbide and a new company called Nanodyne, a joint venture of Rutgers and Procedyne, a local manufacturer of industrial fluidized-bed reactors. Nanodyne has exclusive rights to intellectual property created in Kear's laboratory, and McCandlish is now the company's chief scientist. Nanodyne is one of two U.S. companies starting to make nanophase materials in bulk.

Nanophase materials were until recently made only in laboratory quantities, by evaporating the ingredients in an inert gas and letting them condense on a cool surface. The Rutgers researchers decided to make the powders from water solutions instead. They eventually found cobalt and tungsten compounds that could be mixed and dissolved in water, aerosolized and then spray-dried to produce extremely fine mixtures of tungsten and cobalt salts of any desired composition.

Then McCandlish and Kear perfected a way to make the tungsten react with carbon to yield tungsten carbide. The trick involves passing a mixture of carbon monoxide and hydrogen over the metal salts, an operation they decided could be done in an industrial fluidized-bed reactor, a device in which fine particles are agitated in a stream of gas. The result is particles of tungsten carbide less than 50 nanometers in diameter embedded in larger grains of cobalt.

With the help of local and federal grants, Nanodyne is now producing hundreds of kilograms of nanophase cobalt tungsten carbide every year and sending samples to a number of potential customers for testing. McCandlish says that nanophase cobalt tungsten carbide is substantially harder than the conventional material and that cutting tools made from it appear to wear at about half the rate. Although the nanophase product is likely to be several times the price of its conventional equivalent, its improved properties mean it could capture a substantial fraction of the world market, estimated to be some 20 million kilograms.

Users can plasma-spray the powder onto surfaces to make coatings or press and bake it just enough to make the grains stick together to form a rigid shape. Kear and McCandlish foresee nanophase wire-drawing dies, sealing rings in pumps, bulldozer teeth, industrial knives and drill bits ranging from the very large (those used in well drilling) to the very small (for printed circuit boards).

Maurice L. Gell of Pratt and Whitney, an engine manufacturer, suggested at a recent conference that Nanodyne's materials might find near-term applications in jet engines, and the two companies are now collaborating. Pratt and Whitney is planning a consortium to exploit nanophase materials. Kear also thinks that by borrowing clean-room processing from the electronics industry, it will be possible to produce "flaw-free" critical components such as gears for helicopters.

Nanodyne has one competitor in the brave new world of nanophase materials. Nanophase Technologies in Darien, Ill., is concentrating on ceramics such as silicon nitride, as well as aluminum, zirconium and yttrium oxides. Although conventional ceramics tend to be too brittle for many applications, research by Richard W. Siegel of Argonne National Laboratory, Horst W. Hahn of Rutgers and others has shown that nanophase ceramics can be plastic or even superplastic at high temperatures—that is, they can be substantially deformed without cracking. "That's where there will be commercial applications—you can do pressing and forging into any shape you want," Siegel says.

One problem in making nanophase ceramics, explains John C. Parker, scientific director at Nanophase Technologies, is controlling the size of the particles during the sintering, or baking, part of the manufacturing process. The particles tend to agglomerate and grow, which destroys their properties. Parker says his company has overcome that difficulty. All of the oxide ceramics Nanophase Technologies is developing are transparent. According to Parker, that proves the particles and the pores between them are less than 10 nanometers across, well below the wavelength of visible light.

Parker expects that some of the company's nanophase ceramics will be used in electronic sensors. He has demonstrated that gases can diffuse through the fine pores of the materials and alter their properties in ways that can be detected electrically. And Nanophase Technologies has contracted with Caterpillar in Peoria, Ill., to test ceramic components in vehicles. Parker says the company has already targeted some automotive uses.

Other niche applications for nanomaterials are visible on the horizon. At the

National Institute of Standards and Technology in Gaithersburg, Md., Robert D. Shull is investigating the potential of nanocomposites in magnetic refrigeration equipment. Magnetic refrigeration is now used only within 20 degrees of absolute zero, but new materials could extend that range, Shull says—although he doubts that nanocomposites will ever be cost-effective in conventional refrigerators. Ami E. Berkowitz of the University of California at San Diego recently showed that nanocomposites have magnetic properties that could make them useful in playback heads in tape recorders, and Ronald F. Ziolo of Xerox in Rochester, N.Y., has produced transparent nanophase magnets that might find uses in copiers or facsimile machines.

Parker acknowledges that the nanophase ceramics made by his company are expensive. They are still generally made by condensing metal from vapor on a cold surface and reacting the resultant nanophase powder with oxygen. That technique limits production runs to a few hundred grams at a time and means a price of several hundred dollars a pound. Nanophase ceramic engine blocks will not be used in ordinary automobiles anytime soon.

Even so, the developers do not expect to wait long for real applications. Kear, for one, maintains that his university's direct involvement in the joint venture with Procedyne—a first for the institution—has speeded up the transfer of the new cobalt tungsten carbide to the marketplace, so that commercialization will take from three to eight years rather than the 10 to 15 that are standard. "Nanodyne will be a halfway house for commercialization," he says.

Kear rejects the suggestion that the tight linkage between his laboratory and Nanodyne could stifle free inquiry. "I don't think the science is being corrupted—it's being enriched," he insists. To prove the point, he mentions several other nanophase ideas he has up his sleeve: molybdenum-silicon nanocomposites, which perform well at high temperatures, and other tungsten-based materials. Kear maintains that by passing various gases through unconventional crystalline forms of nanophase tungsten, he can make completely new compounds. "We're doing atomic alloying," he boasts.

Moreover, with the company already shipping samples to potential customers, Kear says that responses from industrial researchers come straight back to his laboratory. A tight connection between laboratory and market may be Nanodyne's competitive advantage today and industry's standard tomorrow. *—Tim Beardsley*

Billions of Buckytubes

Mass production of carbon cylinders sparks interest

This summer, less than a year after NEC Corporation researcher Sumio Iijima displayed the first photomicrographs of nanometer-size carbon cylinders to a crowd of amazed scientists, two of his colleagues have found a way to make the stupendously tough tubules by the gram. Now investors are as interested as the scientists.

Shortly after Thomas W. Ebbesen and P. M. Ajayan described their method in *Nature*, NEC began receiving inquiries. Companies around the world want the tubules because their crystalline perfection makes them stronger than any other material and gives them a potpourri of electrical properties.

The carbon tubes have been dubbed buckytubes, after buckminsterfullerene, the 60-atom spheres that have a structure similar to the late R. Buckminster Fuller's geodesic dome. Buckytubes may be considered as "stretch" fullerenes; in effect, they consist of graphite sheets that have been spliced between two halves of a spherical fullerene. Together the long and the globular species constitute a third form of carbon, after diamond and graphite, that extends the element's reach throughout Euclid's empire. Diamond lives in three dimensions, graphite sheets in two, buckytubes in one and buckyballs in none.

Iijima discovered the tubules more than a year ago while examining the electrodes of a carbon arc such as that used to produce carbon 60 and fullerenes with other numbers of carbon atoms. He found that all the tubules occurred as two or more concentric cylinders. They were as narrow as two nanometers and as long as a micron or more.

Then, this past March, Ebbesen and Ajayan discovered that by altering the pressure of the inert gas surrounding the arc and adjusting other parameters, they could deposit a hard lump of material about five millimeters in diameter on one of the electrodes. When they cracked open the lump, they found a core of nearly pure buckytubes. The total conversion rate was 25 percent—more than 10 times the yield of generators producing spherical fullerenes.

Because of that efficiency, Ebbesen says, buckytubes are inherently cheaper than fullerenes. Moreover, production should be easier to scale up. Because buckytubes form on the rod, not in the arc itself, as fullerenes do, "you can in principle get an even better yield with a larger rod," Ebbesen says.

He and Ajayan have already determined that a mixture of many kinds of tubes conducts electricity fairly well, implying that some of the configurations conduct extremely well. He says NEC scientists are now working hardest on characterizing various configurations. To do that, they must first generate tubes of a given diameter and structure and then measure the electrical conductivity of individual tubes. When asked how they intend to do that, Ebbesen politely declines to answer.

Theoreticians calculate that the most interesting properties will emerge only in the smaller, nanometer-size diameters, where quantum effects become important. "The high-conductivity forms we calculated should be absolutely metallic, potentially as good a conductor as copper," says John W. Mintmire, a theoretical chemist at the Naval Research Laboratory in Washington, D.C. "What we figure happens is that the mobility of electrons compares with that in graphite, but a buckytube has a higher number of charge carriers," he says.

Other buckytube variants should be insulators or semiconductors. But so far none has been shown to function as a superconductor, as fullerenes do when doped, or seeded, with alkali metals. The range of properties heartens electronics engineers because it gives them the flexibility to design extremely small devices. For such devices to become practical, however, a way must be found to make them precisely and in quantity.

What about strength? David

***CONCENTRIC TUBULES** of carbon are so narrow that they conjure up quantum effects. The computer graphic is provided courtesy of NEC/Handa.*

Tománek and his colleagues at Michigan State University calculate that carbon 200, a rather thick, single-walled cylinder, would be "100 times stronger than an iridium beam of the same length." Ebbesen doesn't like to speculate about buckytube girders. "In the short term, what's important is their potential for strengthening composite materials. For that, all you need is a sufficient ratio of length to diameter," he says.

Richard E. Smalley of Rice University, a co-discoverer of the fullerenes, is more inclined to dream. "Now all you've got is lint—to make a shirt you need a continuous fiber on a spool," he says. "We want to do it, and I am confident that NEC is trying to do it. I've spoken to Iijima, and he says that's his single challenge in life." *—Philip E. Ross*

Diagnosis by DNA

DNA-sequencing methods show clinical promise

Just four letters can answer age-old questions about the causes of disease. Strings of *G, C, A* and *T*—the nucleic acid bases guanine, cytosine, adenine and thymine—spell out genes that instruct cells to perform critical functions. Or not. The simple switch of a *T* for an *A* causes sickle cell anemia, for instance. Recently the severity of several major cancers has been linked to changes in a gene called p53.

The process of linking genetic variations to disease is being vastly simplified by machines. Much work on improving the equipment is being done with government funding under the auspices of the international Human Genome Project. Now commercial interest is growing as well. Sequencing technology promises to be a powerful tool for clinical diagnosis. "We want to launch products that drive toward the physician, not lab technicians," announces Jonathan Briggs, acting director of Affymetrics, a newly formed subsidiary of Affymax Research, a biotechnology firm in Palo Alto, Calif.

"Physicians and researchers are using DNA-sequencing technology to develop clinically important tests," observes Michael Hunkapiller, executive vice president of Applied Biosystems, Inc. (ABI), in Foster City, Calif. The firm pioneered the first automated DNA sequencer and is still the lead supplier of systems that break up DNA, attach a telltale fluorescent label to each base and reassemble the sequences. An optical reader feeds the light signals into a computer to record the order of the bases.

Richard A. Gibbs, head of the center for DNA sequencing at Baylor College of Medicine, uses fluorescence sequencing to follow families suspected of carrying one of the more than 100 mutations that cause cystic fibrosis. "Instead of asking simply, does a gene work or not, we're asking, what are the specific differences in this person's DNA?" he explains. Gibbs and others are also using the technique to compare the DNA in different strains of the HIV virus that causes AIDS "to see who got what from whom when." Such glimpses of genetic epidemiology may help in treatment, he notes, by detecting certain bases known to be associated with drug resistance.

The sequencing systems available today, however, still require the artful attention of Ph.D.-level researchers to review the computer's calculations. It can take weeks to edit a sequence such as one from a cystic fibrosis gene. So researchers at ABI, the California Institute of Technology and other institutions are striving to develop algorithms that will enable computers to make reliable diagnoses.

One promising scheme for automating diagnosis is based on computer chips that can directly detect aberrant genes. Developing such DNA chips is the goal of several research teams, including the Houston Advanced Research Center (HARC) in The Woodlands, Tex., and Affymetrics. Researchers pack snippets of DNA that are eight bases long onto the surface of a chip and expose it to a patient's sample. The fresh strand clings, or hybridizes, to complementary bases (every *G* with *C*, *A* with *T*), wending its way in a pattern that reveals whether a patient's DNA has bound to a normal gene or to a mutated sequence laid down on the chip.

"We call our approach geneosensor technology," says Kenneth L. Beattie, head of HARC's DNA Technology Laboratory. The key to his team's diagnostic method is a sensitive detection system that picks up changes in electrical properties when nucleic acid bases bind together. A computer keeps track of the sequence and position of each eight-base probe so that any change in the electrical charge of a given area is quickly translated into a sequence reading.

Researchers expect that they will be able to immobilize arrays of DNA dense enough to screen samples from several patients at once or to do an individual's full genetic workup. "The computational task of diagnosis will become more complicated as chip densities increase," says Mitch Eggers, an electrical engineer responsible for wiring up HARC's chips. But eventually electrical information will not need to be fed into a separate computer, he predicts, because a microprocessor will sit right on the chip.

The development of simple DNA diagnostic techniques will build a market for genetic testing where none exists today, asserts Briggs of Affymetrics. But acceptance will also depend on whether the tests yield simply gloomy predictions or information that can actually guide treatment. Developers of DNA diagnostics realize that the Food and Drug Administration will eventually regulate tests offered for commercial rather than research purposes. Congress, too, is sure to begin paying attention as the public—and health insurance companies—ponder a suddenly clearer vision of the future. *—Deborah Erickson*

KENNETH L. BEATTIE displays a prototype "geneosensor" developed at HARC. Such computer chips bearing myriad bits of DNA may quickly reveal whether an individual carries genes related to disease. Photo: Dan Ford Connolly/Picture Group.

The Business of Education

Do economists have anything to offer U.S. schools? The free market is the latest in a long list of pet nostrums intended to rescue an education system that produces less visible learning per dollar spent than does nearly any other in the world. Virtually every other learned cabal has offered its favorite remedy at one point or another, from social scientists to particle physicists. Occasionally such efforts have led to noteworthy feats—in the mid-1960s Richard P. Feynman read every elementary school textbook proposed for adoption in the state of California.

Economists who are proponents of "educational choice" contend that public schools produce uneducated graduates because the institutions are for the most part monopolistic suppliers. Even if parents send their child to a private school, they must still pay public tuition through taxes. Public schools, free marketers contend, have no profit motive and thus no incentive for effective teaching.

Terry M. Moe and John E. Chubb, two former fellows of the Brookings Institution, have argued further that democratically elected school boards and their bureaucracies stifle the personal initiative that teachers bring to their work. (Moe is now a professor at Stanford University, and Chubb has joined the Edison Project, which aims to set up high-technology, for-profit private schools.)

If the government just gave parents the portion of taxes allocated to education and let them choose how to spend the money, the reasoning goes, schools would have to compete. The ones that provided the best education and attracted the most students would prosper and expand, and the low-quality schools would eventually fold. Overall, proponents argue, the quality of education would improve.

Although many school districts now allow their students to choose which public school they will attend or maintain "magnet" schools intended to attract the brightest and most talented to particular programs, they still generally juggle enrollment so that even the worst-performing, least popular school is not left entirely without students. Devout free-marketers such as Moe and Chubb insist that true educational choice requires not only unfettered selection among public schools but a choice between public and private as well. President George Bush's latest education initiative reserves half a billion dollars for state programs that would use vouchers to help parents send their children to private schools [see "Teaching Real Science," by Tim Beardsley, page 78].

Opponents of such a plan argue that vouchers do not a market make. Instead they would take money from underfunded schools and give it to people who can already afford to opt out of public education. Working from 1980 California census data, Thomas A. Downes of Northwestern University deduced that halving tuition costs would increase the demand for private schools by only about half a percentage point. The data are sketchy, he concedes, but they suggest that parents choosing a school are more sensitive to the perceived quality or the ethnic composition of local public schools than they are to tuition cost.

One fundamental problem with vouchers, argues economist Alan B. Krueger of Princeton University, is that education is not an easily defined commodity. In an efficient market, producers have a good idea of what their customers want, and customers can judge the quality of the product being offered for sale. But currently no one has an unequivocal grasp of who the "customer" is: students, their parents, the firms that will eventually hire them or the society of which they will form a part. And what the customer wants is at least as nebulous. Is it an enjoyable learning experience, a well-rounded education, vocational and professional skills, the values of good citizenship or something else entirely?

Should all schools compete and risk going under if they cannot attract enough students?

Moe has a ready answer to this conundrum: let the marketplace decide. The customers are parents and students with vouchers to spend, and new schools, he predicts, will spring up to offer them whatever they desire. He cites the case of the East Harlem school district in New York City. Abandoning traditional rules freed teachers there to create junior high schools for the performing arts, mathematics, science, environmental studies and other fields. As a result, student performance rose from among the worst in the city to above average.

Moe argues that allowing private schools into the act will provide the same kinds of educational choices more effectively and at less cost than simply mandating public school diversity. The government does not guarantee the survival of private schools, and so they have cost-effectiveness incentives that public ones lack. Furthermore, he says, they are not encumbered by bureaucracy or extracurricular political agendas such as desegregation and bilingual education.

Moe envisions only the most minimal of regulations, perhaps a nationwide set of tests that students must pass for their schools to maintain accreditation. A state would give disadvantaged students larger vouchers so that schools would have an incentive to accept them, and accreditation requirements would ensure that such students were taught effectively, he says.

Krueger warns, however, that private schools could—as other industries once did—turn to style over substance, relying on marketing gimmicks rather than a solid product to bolster their enrollments. Current data, he says, offer little evidence that private school students outperform their public school counterparts once factors such as parents' wealth and education are excluded. Bush's $500 million in vouchers, he notes, is a powerful incentive for the best-looking statistics money can buy.

Krueger is challenging the Edison Project to show that private schools truly can do a better job of educating anyone—not just the well-off or highly motivated. He is suggesting that the new high-tech schools admit 1,000 low-income students by lottery (rather than academic or financial credentials) and track their progress against that of students who entered the lottery but were not admitted. If the country is going to embark on an expensive educational experiment, he says, someone might as well record the results. —*Paul Wallich*

Murder at Ghastleigh Grange

It was evening in 221B Baker Street. Holmes was playing an Irish air on his violin when I was forced to interrupt him.

"Holmes, we just received a letter that describes a most urgent matter."

"Please read it to me, Watson."

"The return address is Ghastleigh Grange at Grimly Sinister:

> Dear Mr. Holmes,
>
> There has been a terrible murder. Miss Melpomene Beetroot has been bludgeoned to death with a chandelier. The police are baffled. Please help us solve this horrifying crime.
>
> Cornelian, Duke of Ghastleigh.

"Watson, we have no time to waste. Pack our bags and hail a hackney to take us to the station. Such a singular crime will test our wits to their utmost, I do not doubt."

Arriving at the nearby village of Grimly Sinister, Holmes and I could see the Ghastleigh Grange. The old castle comprised 46 towers arranged in three concentric circles about a single central tower. They were linked together by narrow walkways that were all several stories high. The only visible entrance lay across a drawbridge to the westernmost tower [*see illustration on this page*].

We were met at the entrance by the butler, who introduced himself as Dunnett. He led us up a winding stairway and along a walkway to a neighboring tower, where we were greeted by the Duke of Ghastleigh.

"Oh, Mr. Holmes," he exclaimed. "I cannot thank you enough."

"No gratitude is required."

"The murder occurred in Miss Beetroot's room. Do you wish to inspect it?"

"In a moment, your Lordship. I must first send a telegram."

"Dunnett will take care of it, Mr. Holmes. We shall wait—it won't take long." Soon the butler returned.

"Please follow me, gentlemen," the duke said. "Each tower of Ghastleigh Grange is but a single large room, and each is inhabited by one of the surviving members of the Ghastleigh lineage." They entered Miss Beetroot's tower.

"This is where the dastardly deed was done," the duke sighed. "There used to be a large chandelier hanging from the center of the roof, but the murderer caused it to become detached in some manner, and poor Melpomene was asleep immediately below it."

"Who found the body?" I asked.

"I did, sir," the butler replied. "What was left of it, sir."

"So you were the last person to have seen Miss Beetroot alive?"

"Apart from the murderer, yes."

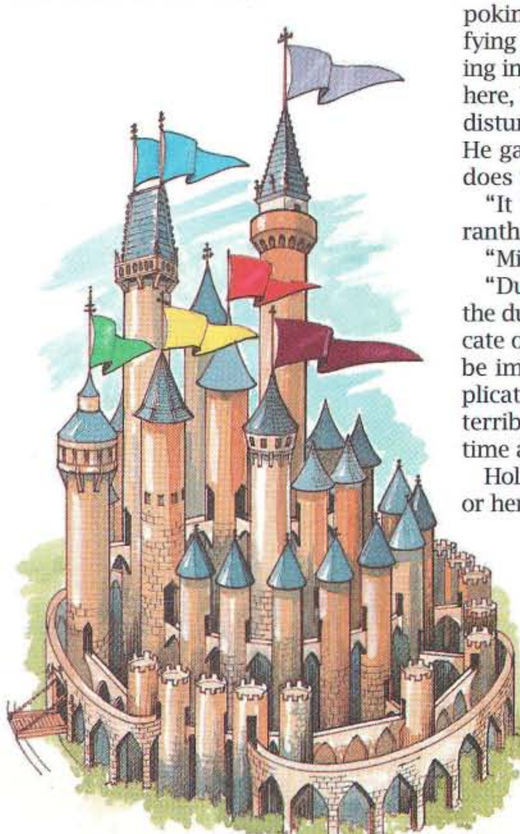
I noticed that Holmes, who had been poking about the room with his magnifying glass, had stopped and was listening intently. "I fear we will find no clues here, Watson," he said. "The police have disturbed the room far too thoroughly." He gazed about him. "To whose tower does this door lead?"

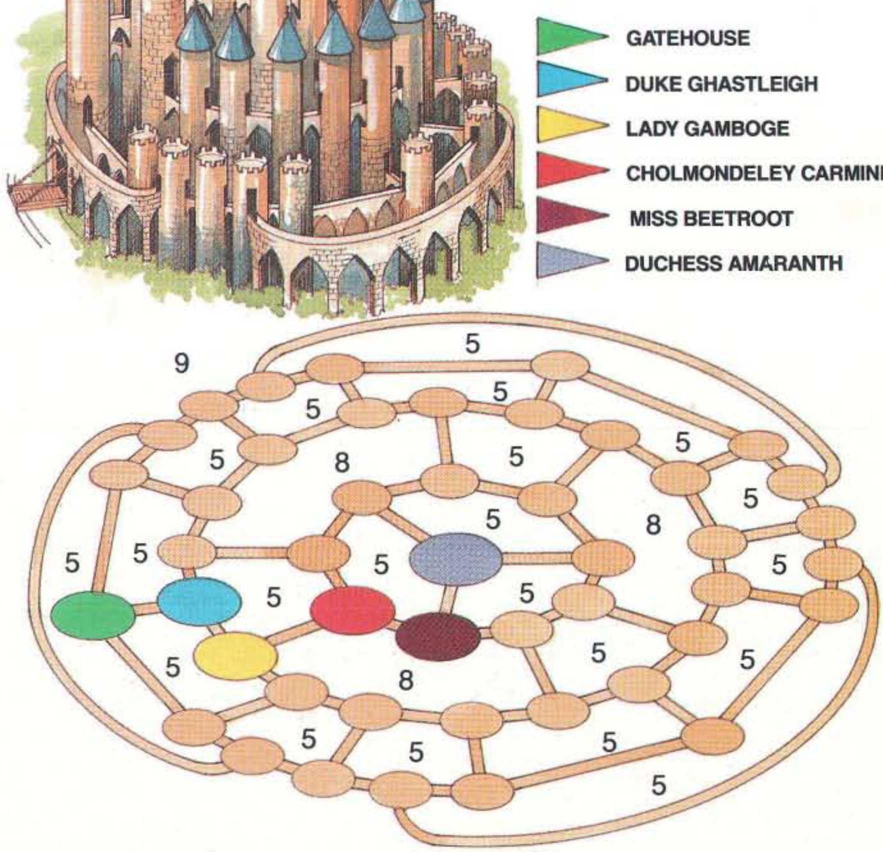
"It belongs to the Duchess of Amaranth," Dunnett said.

"Might she be the murderer?"

"Dunnett has the only set of keys," the duke said. "They are extremely intricate of design, and I do believe it would be impossible for anyone to make duplicates. But in any case, the duchess is terribly deaf and spends much of her time asleep."

Holmes nodded. "Was everybody in his or her room on the night of the crime?"

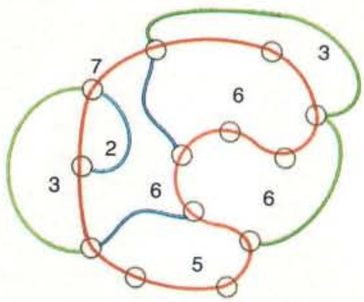




GHASTLEIGH GRANGE is the scene of the crime. Miss Beetroot has been murdered in her room. Numbers show how many walkways surround each region.

The Grinberg Formula

The network below has 13 nodes connected by 19 paths. A closed loop (*red*) visits each node of the network precisely once. Such a loop is called a Hamiltonian circuit. The paths not included in the loop can be described as inside diagonals (*blue*) or outside diagonals (*green*). The loop and the diagonals form several regions, each with a number of surrounding paths, or sides.



If f_j is the the number of regions inside the circuit that have j sides, then $f_2 = 1$, $f_5 = 1$ and $f_6 = 2$. Likewise, if g_j is the number of regions outside the circuit that have j sides, then $g_3 = 2$, $g_6 = 1$ and $g_7 = 1$. Since the network contains a Hamiltonian circuit, it must satisfy the Grinberg formula, which in this case is

$$(f_3 - g_3) + 2(f_4 - g_4) + 3(f_5 - g_5) + 4(f_6 - g_6) + 5(f_7 - g_7) = 0.$$

When the values for f_j and g_j are substituted in the left side of the equation, the result is

$$(0 - 2) + 2(0 - 0) + 3(1 - 0) + 4(2 - 1) + 5(0 - 1),$$

which is indeed equal to 0.

"Almost definitely," the duke sighed. "All my relatives are confined to their rooms at night by the terms of my grandfather's will. The first duke had a terrible fear of loneliness, so his will requires each of his descendants to spend every night at the Grange or else forfeit all claims to the family fortune."

"That's right, sir," Dunnett said. "Every evening I check each tower in turn and lock all the communicating doors. Every morning I make my rounds again and unlock them. On that awful morning, I knocked on Miss Beetroot's door. There was no reply. I became concerned and unlocked her door. That was when I saw the . . . remains."

"And nobody entered or left Ghastleigh Grange during the night?"

"No, sir. Never, sir. The only way to do so is to pass through neighboring towers until one reaches the entrance tower, where I spend every night. I can confirm that nobody passed in or out."

"And all residents were in good health when you locked them up?"

"Yes, sir. They all have bellpulls in their towers, sir, and once the doors are locked they ring through to the duke's tower to confirm they are present."

"I use the bells to check that they comply with the terms of the will," the duke said. "Dunnett is correct. I keep a strict log, and every bell was rung."

"But an intruder could easily ring a bell," I protested.

"No, Dr. Watson," the duke replied. "Each resident has a personal code, known only to him or her and myself."

Holmes turned to Dunnett. "Did you enter any tower more than once?"

"Oh no, Mr. Holmes," the butler cried. "I enter a tower only once when I make my rounds—that's an inviolable rule. It would never do to disturb anyone once his or her tower has been locked."

Holmes tried another line of inquiry. "Your Lordship, did the police discover the estimated time of death?"

"It was impossible to tell, Mr. Holmes, because of the state of the body. From the degree to which the blood had congealed, they thought it was probably before midnight."

Holmes's brow furrowed. "Is there any way a person could pass between the towers except along the elevated walkways?"

"A skilled mountaineer might perhaps scale the walls from ground level. But, Mr. Holmes, he would never be able to do so at night. The Ghastleigh family is very conscious of security. Dunnett's last act on his nightly rounds is to release a pack of hounds onto the grounds of the Grange."

A boy from the post office arrived at the entrance with a telegram for Holmes—no doubt a reply to his earlier missive. As he read it, I saw his eyes narrow. "What route do you take on your rounds, Dunnett?"

"It varies, sir."

"Can you recall the route you took on the evening of the murder?"

"No, sir."

"That is most unfortunate." Holmes shook his head sadly. "Watson, we shall hire rooms there for the night. We can do no more here."

"But, Mr. Holmes, the murder—"

"I did not say, your Lordship, that I cannot solve the crime. I was merely remarking that my investigation here is complete. Dr. Watson and I have some work to do, and I have no doubt whatsoever that shortly afterward I will name the criminal. Dunnett, please call a carriage."

Holmes and I retired to a comfortable inn in the village of Grimly Sinister. "Holmes, did you really mean what you said to the duke? About being confident of naming the murderer?"

"Watson, since when have I ever dissembled to a duke?"

"But—there is so little to go on."

"Nonsense, Watson! Let us rehearse the pertinent facts. Miss Beetroot was killed before midnight. No intruders could have gotten in or out because of the hounds. The murderer was therefore one of the inhabitants of Ghastleigh Grange. Dunnett locked the inhabitants in their separate rooms, and they reported their presence to the duke once he had done so. Dunnett started from the entrance tower and returned to it, entering each tower precisely once. After the rooms were locked, the only residents who could have entered Miss Beetroot's room without being observed were her immediate neighbors. But to do so, they must have had a key, and Dunnett possesses the only set of keys. The keys cannot be copied. Who, then, could have committed the murder?"

"Uh—oh, of course! Dunnett! He could have returned to the room after Miss Beetroot had rung her bell to say she was present."

"Precisely. The duke told us that the Duchess of Amaranth, whose room is adjacent, is as deaf as a post and sleeps heavily. Dunnett could have waited in the duchess's room until Miss Beetroot had rung through and then reentered to kill her."

"With a chandelier?"

"He killed her with another weapon—a length of lead pipe, perhaps—and then arranged for the chandelier to fall to obliterate the evidence."

"An interesting theory," I told him.

He nodded. "But as yet only a theory, Watson. How can we prove Dunnett was the killer? The Duchess of Amaranth would still be asleep when he reemerged, and then he would continue his rounds as if nothing had happened."

"The crash of the chandelier would surely have awakened somebody—"

"The towers are isolated from one another. No, there would be no sounds."

"Dunnett would be delayed on his rounds."

"A few minutes only, if he had already weakened the chandelier's support. Not enough to attract attention."

I struck my forehead a blow with my hand. "Then we are defeated, Holmes! It can be none other than Dunnett, yet the scoundrel must go free."

Holmes laughed. "By no means, Watson. If fortune is with us, he may yet be condemned out of his own mouth." He handed me a sheet of paper. On it was a map of Ghastleigh Grange. "Watson," he said, "here is a simple puzzle for you. Dunnett claims that his evening rounds took him through each tower once and once only. He cannot have moved from one tower to the next except along the walkways. Perhaps you will find such a route for me."

"Certainly, Holmes, there must be hundreds of solutions."

"In truth, Watson, I suspect that there is not even one. I have asked you to find a Hamiltonian circuit: a closed loop through a network that visits each node precisely once. It is named for Sir William Rowan Hamilton, who marketed a puzzle asking for just such a circuit along the edges of an icosahedron. No foolproof method is known, beyond systematic trial and error, that can determine whether a given network has a Hamiltonian circuit."

"Then Dunnett has cheated the noose, Holmes, for the network is too large to be analyzed by trial and error."

"Not necessarily, Watson. I hope that in this case we may be fortunate. I recently read about the fascinating work of a Russian mathematician by the name of E. J. Grinberg. He has devised a condition that must be satisfied by any network in the plane that possesses a Hamiltonian circuit. We will see whether it is satisfied by the towers of the Grange. Please check my work, Watson."

"I'll do my best, Holmes."

"Imagine a network that can be drawn in the plane—that is most important, Watson. Although the walkways of Ghastleigh Grange are elevated, they do not cross one another. Let such a network consist of n nodes and some number of paths connecting those nodes [*see box on opposite page*]. Take for granted that a Hamiltonian circuit connects the nodes—"

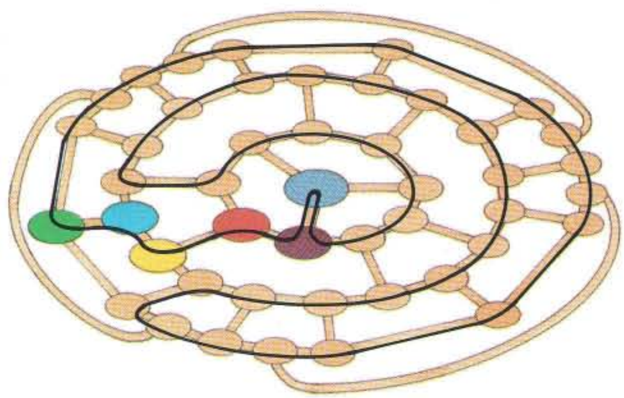
"In other words," I interrupted, "you mean we're assuming that there is a closed path through the network that goes through each node only once."

"Exactly. If that's the case, we can deduce certain characteristics of the network. First of all, we know that the Hamiltonian circuit has n edges since it visits each node once. The connections that do not belong to the circuit run 'diagonally,' either across the interior or the exterior of that circuit. The interior is divided into a certain number of regions by those diagonals that cross it. If there are d diagonals, the number of regions must equal $d + 1$."

"Why is that, Holmes?"

"Imagine adding the diagonals one at a time. The Hamiltonian circuit itself borders one region, and each diagonal produces an additional region. Now, there is an alternative way to count the number of regions. Each region has a certain number of sides—edges of the network that surround it. Suppose that for each integer j, f_j is the number of regions that have j sides. Then the total number of regions in the interior is also given by $f_2 + f_3 + \ldots + f_n$. And therefore $f_2 + f_3 + \ldots + f_n = d + 1$."

"Many of these f_j are just equal to zero. Am I right?"



DUNNETT cannot visit all the towers of Ghastleigh Grange without retracing his steps. As shown above, he must reenter at least one of the towers.

"Yes. Next, I count the number of edges that surround these regions in two different ways. Any region with j sides is bounded by j edges, so such regions contribute jf_j to the total."

"So, Holmes, the total would be $2f_2 + 3f_3 + \ldots + nf_n$."

"Well, not quite. In such a count, each of the d diagonals is counted twice, once for each of the two regions that it abuts, but the n edges of the circuit are counted only once. Therefore, $2f_2 + 3f_3 + \ldots + nf_n = 2d + n$. Now, doubling the first equation and subtracting it from the second, I find that

$$f_3 + 2f_4 + 3f_5 + \ldots + (n-2)f_n = n - 2.$$

There is a similar equation for the exterior of the circuit,

$$g_3 + 2g_4 + 3g_5 + \ldots + (n-2)g_n = n - 2,$$

where g_j is the number of regions outside the circuit that have j sides. Finally, subtracting one equation from the other, I attain my goal, the Grinberg formula:

$$(f_3 - g_3) + 2(f_4 - g_4) + 3(f_5 - g_5) + \ldots + (n-2)(f_n - g_n) = 0.$$

"Elegant, Holmes, and assuredly subtle, but I fail to see the relevance to Dunnett's guilt. We have no idea what the values of f_j and g_j should be. Indeed, if there is no Hamiltonian circuit, there are no such values at all."

"I am hoping that the assumption of a circuit will lead to a logical contradiction. Now, Watson, if you examine the network for Ghastleigh Grange, you will find that all of its regions have either five, eight or nine sides. If it has a Hamiltonian circuit, then by Grinberg's formula

$$3(f_5 - g_5) + 6(f_8 - g_8) + 7(f_9 - g_9) = 0.$$

"But there is only one nine-sided region—the entire outside of the network—so that $f_9 - g_9$ is either 1 or –1. Yes, I think we have the wretch. For now we are logically forced to conclude that $3(f_5 - g_5) + 6(f_8 - g_8) = \pm 7$. But it is impossible to find a solution to such an equation since f_5, g_5, f_8 and g_8 are integers. The left side of the equation yields only multiples of 3, and the right side can only equal +7 or –7."

"Therefore, no Hamiltonian circuit exists! Dunnett is lying. Holmes, I am speechless with admiration."

He smiled at the compliment. "Thank you, Watson. Dunnett must have visited at least one tower twice. The only reason he should lie is that it was Miss Beetroot's tower to which he returned. And indeed, at least one possible route visits all the towers—except Miss Beetroot's—only once. Tomorrow we shall confront Dunnett with the evidence of his mendacity."

"Brilliant, Holmes. But what put you on to him?"

"I dispatched a discreet telegram to Scotland Yard to check all of the suspects' files, and I found that his name was Hugh."

I scratched my head, baffled. "How did that help, Holmes?"

"What else could I have concluded after receiving a telegram from Scotland Yard telling me 'Hugh Dunnett'?"

FURTHER READING

GRAPH THEORY. Edited by Ron Gould. Benjamin-Cummings, 1988.

BOOK REVIEWS by Philip Morrison

Before and After Columbus

THE TAINOS: RISE AND DECLINE OF THE PEOPLE WHO GREETED COLUMBUS, by Irving Rouse. Yale University Press, 1992 ($25). **BLUE CORN AND CHOCOLATE,** by Elisabeth Rozin. Paper cut-out illustrations by Barbara Balch. Alfred A. Knopf, 1992 ($23).

The 1,800-mile-long island chain that parts the tropical blue Caribbean from the wide Atlantic bends toward the American continents like a forked sapling in a storm, its southernmost roots set almost at the mainland delta of the Orinoco. Florida is loosely held between the two forked ends. The Gulf-side limb is formed by the four large mountainous islands of the chain, but the ocean-facing one is made by a thousand tiny coral-built islands and cays, the Bahamas.

Two moonlit hours past midnight on Friday, October 12, 1492 (we mark the day by the Old Style calendar: actual day count reckons the modern anniversary as October 21), the Captain-General in *Santa María* heard a cannon shot from his lead ship, fast-sailing *Pinta.* That was the prearranged signal for landfall after their long voyage west. (Just such a hopeful alarm under moonlight had been fired, but in error, from *Niña* only the Sunday before.) This time Columbus was confident, for he felt he had seen a night-light he held to be "a sign from God." His landfall was on a coral island among the Bahamas, though on just which low cay he first set foot remains uncertain within 100 miles or so. There his ships lay to, six miles out at sea, until "at dawn we saw naked people and...went ashore" by boat.

Two worlds of humans, one Old, one New, were stitched together that day as twice before. Five hundred years earlier a few Norse seafarers had made a homestead transiently in Newfoundland. Some 10,000 years earlier still, bands of men and women from northeastern Asia foraged across the lowlands where now lies the cold Bering Sea to populate two continents before the glaciers receded to flood the pioneer paths. (Earlier joinings, and even other routes, are not yet excluded.)

The naked islanders Columbus saw were Tainos. He describes them as "very well-built...tall...none of them has a paunch.... They are friendly and well-dispositioned...they bare no arms save for small spears, and they have no iron." They held alluvial gold, which they beat into small plates for ornament; the chiefs wore cast pieces of gold alloy traded up from South America.

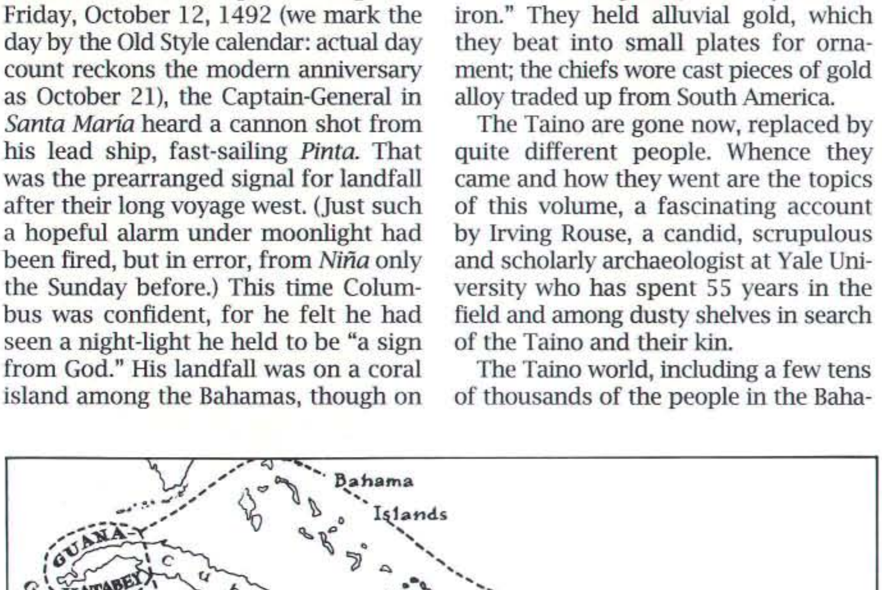
The Taino are gone now, replaced by quite different people. Whence they came and how they went are the topics of this volume, a fascinating account by Irving Rouse, a candid, scrupulous and scholarly archaeologist at Yale University who has spent 55 years in the field and among dusty shelves in search of the Taino and their kin.

The Taino world, including a few tens of thousands of the people in the Bahamas, centered on the four larger islands—Cuba, Jamaica, Hispaniola and Puerto Rico. At both ends of the chain lay the lands of people of other tongues. In western Cuba the Guanahatabeys, and in the southernmost of the Lesser Antilles the Island-Caribs. The Tainos themselves came to number perhaps as many as two million, a thousand or two large villages of wood and thatch, chiefdoms under their caciques, who could be men or women. They lived by a sophisticated tropical agriculture, cropping rows of mounded beds of the soft soils where root crops grew, their staples cassava and the sweet potato. Maize, eaten off the cob, was less important for them than for the mainland peoples. They grew also squash, beans, pineapple, peppers, peanuts, cotton and tobacco. They freely took fish, iguanas, parrots, but of land mammals they knew only dogs. They speared rivermouth manatees for meat.

They were skillful coastal seafarers in paddled dugout canoes, including some boats of giant size that held 150 persons; they had no sails nor blue-water navigation skills beyond expert pilotage. They "traveled by sea whenever possible." So much were they at home on the sea that Taino subcultures can be assigned to specific ocean passages between islands: the two sides of each sea passage tend to be more alike culturally than are groups separated by equal stretches of land.

The West Indies were first settled about 4000 B.C., perhaps by river valley people who worked their tools in flaked flint. Most likely they came by canoe from Belize to Cuba. These were people oriented not to the sea, but to the land. Land mammals, such as ground sloths, were there as game on the largest islands as late as 800 B.C. These early comers did not enter the smaller islands.

The Archaic people, coastal folk from South America, arrived in the islands about 2000 B.C.; their tools were ground of stone, bone and shell. The two groups met first at the Mona Passage between Puerto Rico and Hispaniola. Much later a new distinct group entered from the Orinoco Valley, probably following the strong outpouring currents of that river to the nearest islands. They made pottery, whose shards supply much evidence; they were the forebears of the Taino. A complex set of interactions



AT TIME OF COLUMBUS'S VOYAGE, the Taino inhabited the four larger islands of the Bahamas. At either end of the chain lived people who spoke other languages.

along shifting frontiers finally led to the end of the initial inhabitants. They were replaced by the group who became fully Taino by about A.D. 1200 in a little-documented maturing of that culture across all the central islands. That was the first of the two repeoplings of the island chain.

The signs of mounded agriculture, courts for the ceremonial ball game played with rubber balls from the mainland, some Taino language, some Taino art, and some Taino genes remain in the West Indies. So cavalier a summation cannot do justice to Professor Rouse's discriminating arguments. His is a wise and powerful method, sometimes uniting, sometimes contrasting, the biological, linguistic, cultural and finally the ethnohistoric lines of evidence, based on study of the Spanish chroniclers.

But the Tainos were soon ended as a people. Columbus and his brother brought out several thousand Spanish colonists. Neither their seed crops, their goldfields nor their livestock worked well. The Taino, forcibly taxed for gold, cotton or food, could not pay; they mounted revolts by 1495. Columbus himself, a failed ruler, was sent back to Spain in chains in 1500 for a year or so.

The tales both of circumstance and maltreatment are gloomy; smallpox, military defeat, atrocious punishments, even massacre, famine arising from intolerable demand for forced labor, and overwork steadily and swiftly reduced the Taino population. Taino sent to Spain as slaves proved poor workers, fell ill and were sent back by the crown. Some 40,000 of the "well-intentioned" Taino were removed as slaves. Others were rented to colonists for work in the goldfields (which failed around 1520). In 1515 sugarcane arrived from the Canaries, and the forced labor was set to tend the cane fields and the livestock. By 1524 there were more mainland Indian and black African slaves than Taino in the islands; by 1540 few Taino remained. The Spanish too left, abandoning many towns, drawn by the spoils of mainland conquest. The West Indies became a backwater until the 17th-century rise of the trade in sugar.

Many Taino women had married Spanish settlers, so that a large proportion of the modern population of the Spanish-speaking islands is able to claim partial descent from the Taino. Taino culture appeals to feelings of cultural identity, and across Puerto Rico there are parks and museums that attract the local people through a mix of Spanish, African and Taino traits. The reconstructed dance and ball courts at Caguana, shown in the book in a good

aerial view, offer an outstanding example. A revival of Taino art, some of it of the highest quality, has taken place. A haunting and intricate old carving in shell of the dog deity, from Antigua but of uncertain origin, decorates the book, and we see a number of the enigmatic carved "three-pointers," perhaps the supreme deity, who held the spirit of cassava.

In the rest of the polyglot West Indies, Taino depopulation preceded European colonization; those islanders boast no Taino genes but trace their origins back to Africa, Asia or Europe. European powers first came to challenge Spain once sugarcane became a source of profit about 1625; Spanish-, English-, French- and Dutch-speaking islands were and remain much ruled by the tyranny of sugar.

The population of these twice-repeopled islands has reached now a dozen times the count of the Taino. But the slaughter bench of history, and of prehistory as well, must sadden every recollection of those tragic junctures. There is of course another story to tell—and it is brought to life in Elisabeth Rozin's fine cookbook. On all the continents, every day, many still enjoy what Taino culture gave its sail-borne visitors: cassava, tobacco, maize (in yellow, white, red, black, blue and speckled), chile peppers, squash, pumpkins, beans never before seen, fragrant pineapple, and peanuts. True, the glistening gifts of tomato, potato, chocolate, vanilla and the domesticated turkey, as well as rubber and more, went to the outside world straight from the two continents and not by way of the islands. That initial island list is luminous enough.

With the recipes are photographs, drawings (one shows the genuine Aztec turkey), colorful decorations and accurate, pointed little essays on relevant issues: how corn and beans and alkaline ash are related to good human nutrition, the nonfattening virtues of the unbuttered potato, how to draw on Asia and Africa to confer higher status on the humble peanut. Here are 175 varied recipes that score the products of the gifted people of the New World in a long theme whose variations are as unexpected as purple Peruvian potatoes and blue Southwestern corn, turkey in red chile sauce and beef in a spicy, unsweetened, chocolaty one.

"We repaid them," Rozin writes, "by wiping them out.... But they are...part of us, those ancient farmers and hunters and naturalists and cooks, and their food too is our food.... In eating, if nowhere else, we celebrate the unity that comes from diversity." Let the anniversary proceed in that spirit.

Telescope Territory

UNUSUAL TELESCOPES, by Peter L. Manly. Cambridge University Press, 1991 ($39.95).

A Zen tale recounts how an acolyte was once told to polish a brick into a mirror. The feat remains a challenge, but this delightful little book surveys some near approaches: astronomical telescopes that use mirrors of obsidian, glazed earthenware, specialized ceramics, foamed glass, plastics, rotating pools of mercury, aluminized films pulled concave by vacuum, and cast metal alloys like that Isaac Newton ground and polished. Tokyo Observatory uses a zenith telescope with a porcelain tube to provide high thermal stability. Wooden tubes are shown—and no-tube designs. A fine design exists for a solid Schmidt telescope with the fastest optics ever planned, at f/0.2. It would require a few tons of diamond if built at competitive scale; no date has been set for construction.

Some of these contrivances are inspired amateur improvisation, some are "blatantly odd," others represent the leading edge of professional technology, like the contemporary computer-controlled rapid movement of parts of the optical train to compensate for atmospheric twinkle. Telescope territory is a technical wonderland, the author, himself a wanderer there, nicely reminds us. It is better understood once you recognize that "the basic idea of a reflecting telescope" is to hold at most a few grams of aluminum in a very precise place. The mounting that does the job may weigh up to hundreds of tons; there is clearly room, and to spare, for novel designs.

The smallest telescopes here are not astronomical. Made to help certain cases of defective vision, they are Galilean telescopes that do not invert the image, like one tube of an opera glass. A little telescope one centimeter long, with a clear aperture of three millimeters diameter, is mounted in each lens of a pair of spectacles. The user sees an image magnified 2.5 times near the center of his visual field. Some people can drive a car only when using such an aid. The device amazingly recaps the prehistory of the telescope; years before the Dutch opticians and Galileo, similar small eye telescopes were probably tried out as visual aids by G. B. Della Porta in Naples. To set off such professional elegance, we see a beer can telescope, eyepiece coming out of a mounted can within which a cosmetics mirror forms the image.

Big telescopes are headed by the familiar grand 1,000-foot radio dish in Arecibo, Puerto Rico. Much less known are the biggest refracting telescopes. The working lens at Yerkes Observatory is the largest one of real and lasting utility. Then there is the Great Treptow Refractor in Berlin, built in 1896. It holds a 27-inch lens out at the end of a giant 60-foot tube, longer than Yerkes, and "a massive example of 'iron and rivet' engineering" that weighs 120 metric tons. It was built for public educational use. Observers stand quite still at the center of rotation, and the great machine swings around them. The jacket cover shows a cutaway view of this wonder. A computer-restored engineering drawing to which Susan Manly has added tongue-in-cheek a couple of viewers, only the lady a cynosure, in vermilion period hat and gown with top-hatted escort. (Save the dust jacket!)

There are 150 designs treated here, telescopes of all sorts. Examine the immobile scope, vertical at the South Pole, computer-controlled via satellite link by its creators in Florida, modern savants who know how to live. This cheerful and original book is amusing and genuinely instructive at once.

The Culture of Childhood

I SAW ESAU: THE SCHOOLCHILD'S POCKET BOOK, edited by Iona and Peter Opie. Illustrated by Maurice Sendak. Candlewick Press, 1992 (2067 Massachusetts Avenue, Cambridge MA 02140) ($19.95).

For nearly four decades, the Opies acted as intrepid ethnographers among the great tribe of children—watching, listening, sometimes advancing in giant steps, along the mean streets and open greens of Britain and beyond. In volume after volume, they assembled and annotated the folklore and folkways that pass from child to child in the old and vigorous culture all of us shared for years before we were compelled to depart. Here is their first work of all, reissued, a chunky little book rather bigger than your hand. It is the accumulation of 174 "traditional rhymes of youth," first published in a small edition in the paper-starved postwar year of 1947. Walker Books and their American friends have made us a present of the work anew, including modified endnotes by Iona Opie. They have persuaded a most gifted artist to illuminate the book both in fact and in metaphor with a drawing or two on every page, in colors, pen and pencil.

In the counting rhyme, "I one my mother./I two my mother./I three my mother..../I ate my mother," Sendak shows her inside the voracious baby. But later, in "Rain, rain, go away," she turns herself into a tree to shelter him.

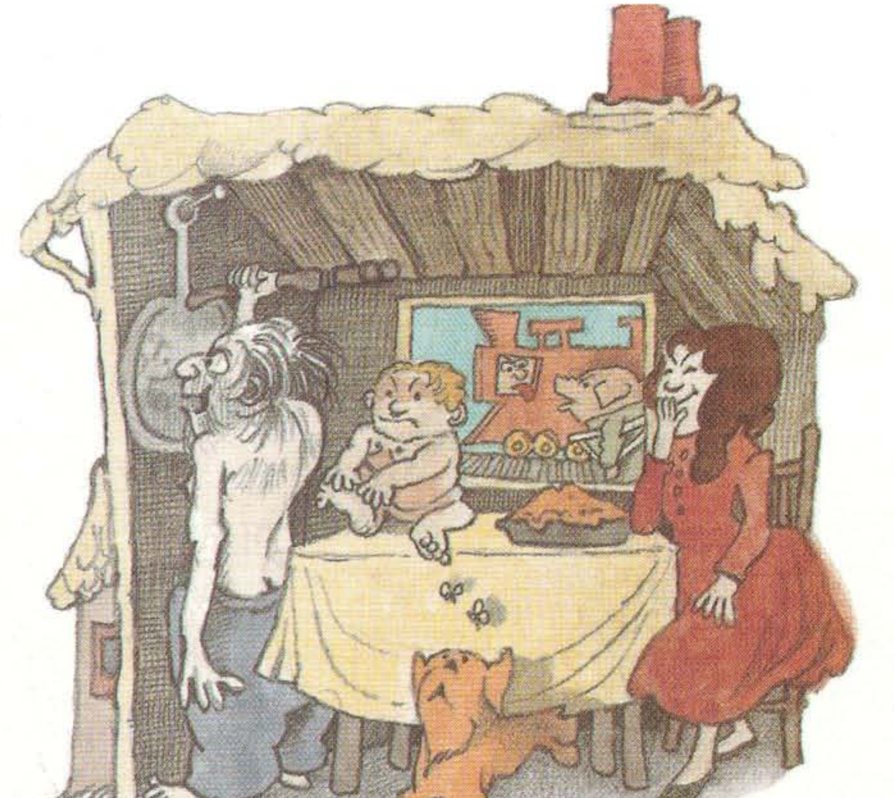
The book presents artful texts, mainly old, irrefutable evidence. In the wordsong, you hear the origin of poetry, the catch of the sound dominant over mere aptness or reason. In the jest and wordplay, you glimpse the rise of wit and humor. Open or covert intention in many of these verses foreshadows our seven deadly sins. (Only avarice, perhaps, is poorly represented among the young; envy, lust, greed, anger, pride and sloth abound.) Here passim you join the popular feast of laughter, "best antidote to the anxieties and disasters of life," to cite Iona Opie's sage introduction.

Sendak certainly sees the joy and power of children but equally reminds them of inevitable suffering. One Satanic drawing repeats Sendak's own exasperating Pierre, who just didn't care but was "boiled till he was done."

An insult: "Tommy Johnson is no good,/ Chop him up for firewood;/ When he's dead, boil his head,/ Make it into gingerbread." A grace: "Bless the meat,/ Damn the skin./ Open your mouth/ And cram it in." (Sendak has painted the old Cramit Inn on the facing page.) Repartee for small folk: "'Have you got a sister?'/ 'The beggarman kissed her!'/ 'Have you got a brother?'/ 'He's made of India rubber!'/ 'Have you got a baby?'/ 'It's made of bread and gravy!'" (You can plainly see that uncommon infant right on the tray.) Latinity: "Brutus adsum jam forte,/ Caesar aderat./ Brutus sic in omnibus,/ Caesar sic inat." An adolescent lullaby: "Good night, sleep tight./ Don't let the bugs bite;/ If they do, don't squall,/ Take a spoon and eat them all."

These have putative sources, some lost in the depths of time, some innocently rediscovered and held as new. "Mairzy Doats," a swing lyric of 1943, was naively claimed original, but a 15th-century verse is cited here: "Is thy pott enty, Colelent?/ Is gote eate yvy./ Mare eate ootys." The locutions are about as old as the medium of transmission; here the mean age is that of the widespread printed word that records Modern English texts, a couple of centuries.

Close with the title verse, a small drama of love and grammar: "I saw Esau kissing Kate,/ The fact is we all three saw;/ For I saw him,/ And he saw me,/ And she saw I saw Esau."



MAURICE SENDAK'S DRAWINGS illustrate Iona and Peter Opie's collection of children's rhymes. Here, "Sam, Sam, dirty old man,/ Washed his face in a frying pan,/ Combed his hair with a leg of a chair."

ESSAY by Gerard Piel

Agenda 21: Sustainable Development

Now that the chiefs of 178 missions have entered their opening statements in the minutes, the 47th General Assembly of the United Nations, convened at its headquarters in New York City, has got down to the real business at hand. This business is Agenda 21, the work product of the U.N. Conference on Environment and Development, held at Rio de Janeiro in the first two weeks of June. The chances are good—that is, if the chances are good for the human future—that Agenda 21 will supply the agenda of many general assemblies to come.

In 40 chapters, Agenda 21 spells out the tasks necessary to secure "sustainable development." It is a program for husbanding the planet's wasting resources. Along with the roster of problems familiar to environmentalists—the ozone layer, global warming, deforestation, desertification, soil erosion, biodiversity—Agenda 21 addresses action to be taken against poverty, infant mortality, malnutrition, epidemic disease, illiteracy and other afflictions that waste that other resource of the planet: its human population.

With the action prescribed by Agenda 21 concerted by the U.N., the earth may be made to support the inevitable increase of the human population to not less than 10 billion toward the end of the next century—and to sustain the population thereafter, stabilized around the number at which it will then have arrived. From chapter to chapter, Agenda 21 prices out the capital requirement for each of the jobs to be done. To each of two classes of nation—the developed and the developing—it allocates their share of the capital necessary to accomplish that task. Allocations to the two parties vary from chapter to chapter, depending on the nature of the required input of technology.

All told, the requirements build up to \$600 billion per year, to be maintained until development becomes self-sustaining. For the developing nations, the cumulative allocations add up to more than three quarters of the annual investment, principally in the form of labor and resources. For the developed countries, that leaves a balance of \$125 billion, to be supplied principally in the form of essential technologies.

The adoption of Agenda 21 by the Rio conference commits no nation to do anything. It carries even less commitment than the two environmental conventions signed by 110 heads of state and government at the Earth Summit that climaxed the proceedings at Rio. The global warming convention was reduced by objections from the U.S. to little more than a declaration of intent. From the biodiversity convention the president of the United States withheld his crucial signature. Yet Agenda 21 will carry more weight than solemn undertakings of the community of nations, or they too will come to nothing.

Agenda 21 concerns nothing less than the accommodation of the human species to the resources of the earth. Logistically at least, it shows this goal is feasible and finite.

In the developed countries, industrial revolution has so increased individuals' material well-being as to bring their population growth to a halt. Assured of the survival of their first infants, people make the decision to stop having children. The celebrated population explosion of the developing countries evidences the beginning of industrial revolution there. The more portable technologies—popular education, sanitation, preventive medicine and the green revolution—have brought lengthening of life expectancy everywhere.

The rate of increase of the world population has been in decline since it peaked at about 2 percent in 1970. In some countries, most notably India and China, birth rates have been declining as the death rates of the under-five-year-olds have fallen.

In its favor, Agenda 21 has a 50-year history at the center of international politics. In the aftermath of World War II and at the founding of the U.N., security meant economic development. With Point 4 in his 1948 inaugural address, Harry S. Truman rallied his countrymen to carry the Marshall Plan, which had begun the rebuilding of Europe, to the underdeveloped countries. For the U.N. General Assembly, a group of experts estimated in 1951 that annual investment on concessional terms of 1 percent of the gross domestic product (GDP) of the developed countries, sustained to the end of the century, would see the underdeveloped countries into their industrial revolutions.

In 1961, even after the cold war had made arms control the synonym of security, John F. Kennedy moved the U.N. General Assembly to declare the 1960s the Decade of Development and committed 1 percent of his country's GDP to the vision. When hindsight could see the 1960s as the Decade of Disappointment, the industrial nations pledged the "more realistic" figure of 0.7 percent of their combined GDP to the economic development of the underdeveloped countries. Nothing, however, came of these unilateral promises from above.

In the 1970s the underdeveloped countries were emboldened by the OPEC oil-price shocks to call for a New International Economic Order. The developed nations not only were to meet their promises of economic assistance, they also were to revise the terms of trade under which the underpriced resources of the preindustrial countries, beginning with oil, subsidized the prolonged expansion of the industrial economies. Nothing came, either, of the unilateral challenge from below.

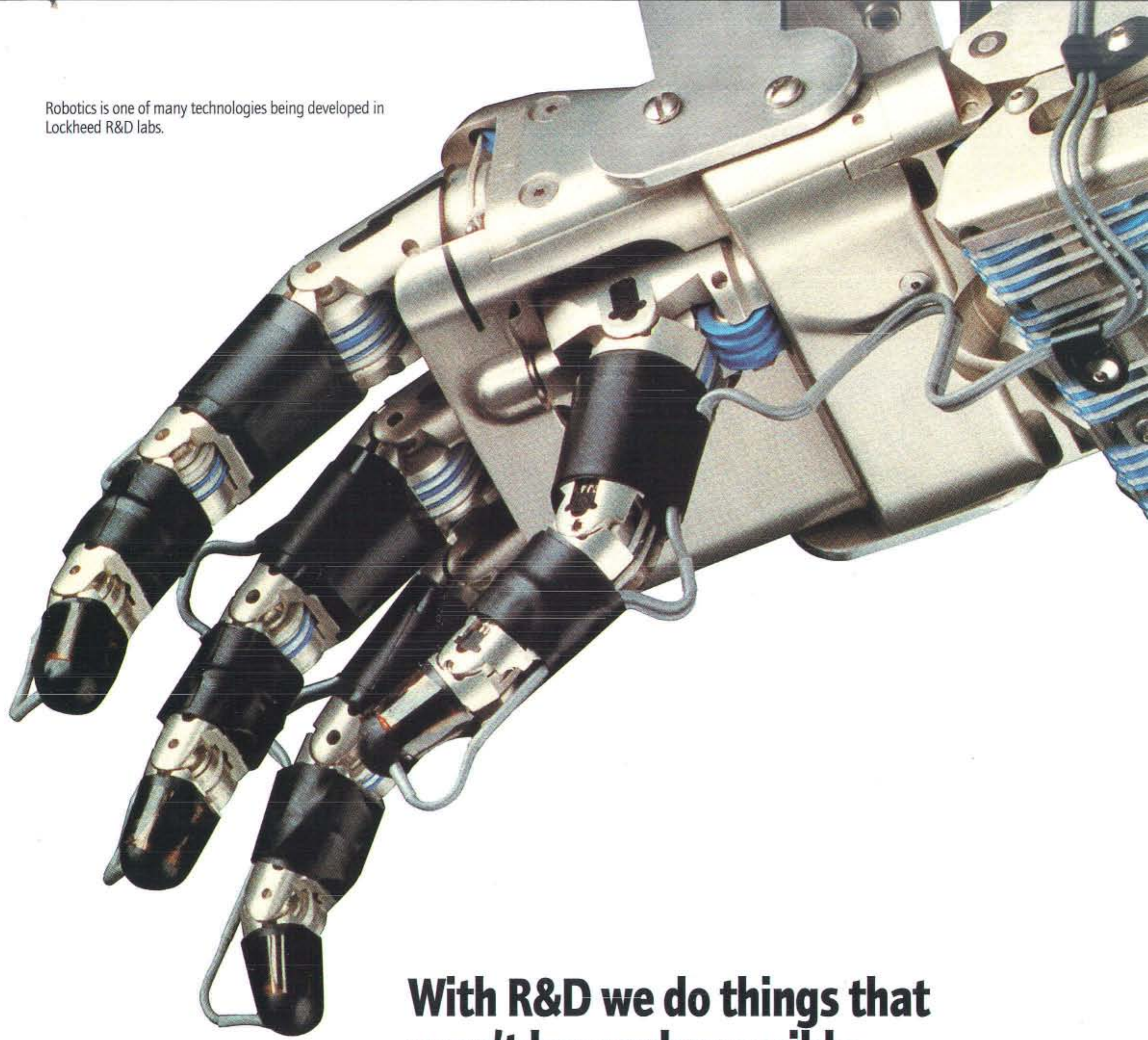
In the framing of Agenda 21, the nations were at least driven to take up the long-overdue task of development by fear, rising worldwide, for the environment. Old habits die hard, however; Agenda 21 is the product of hard bargaining. The 500-page draft went to Rio with all "implementation" (funding) passages bracketed for negotiation there. From Rio, Agenda 21 came inflated to 1,000 pages by the fudging compromises that removed the brackets.

Whatever those compromises, the rich and poor nations have agreed for the first time on what it will take to bring the human species to adjust its numbers and its appetites to the dimensions of a mere planet. They have described the task to be done.

The unemployed people-power and the underutilized resources of the poor nations await the technology to put them to work. The \$125 billion for technology comes to just under the long-promised 0.7 percent of the GDP of the industrial countries.

GERARD PIEL is chairman emeritus of Scientific American. *His most recent book,* Only One World, *addresses the challenge of achieving economic growth in a sustaining environment.*

Robotics is one of many technologies being developed in Lockheed R&D labs.



With R&D we do things that aren't humanly possible.

At Lockheed, our commitment to research and development leads to remarkable innovation. Technologies that once seemed inconceivable are now within reach in our R&D labs across the country.

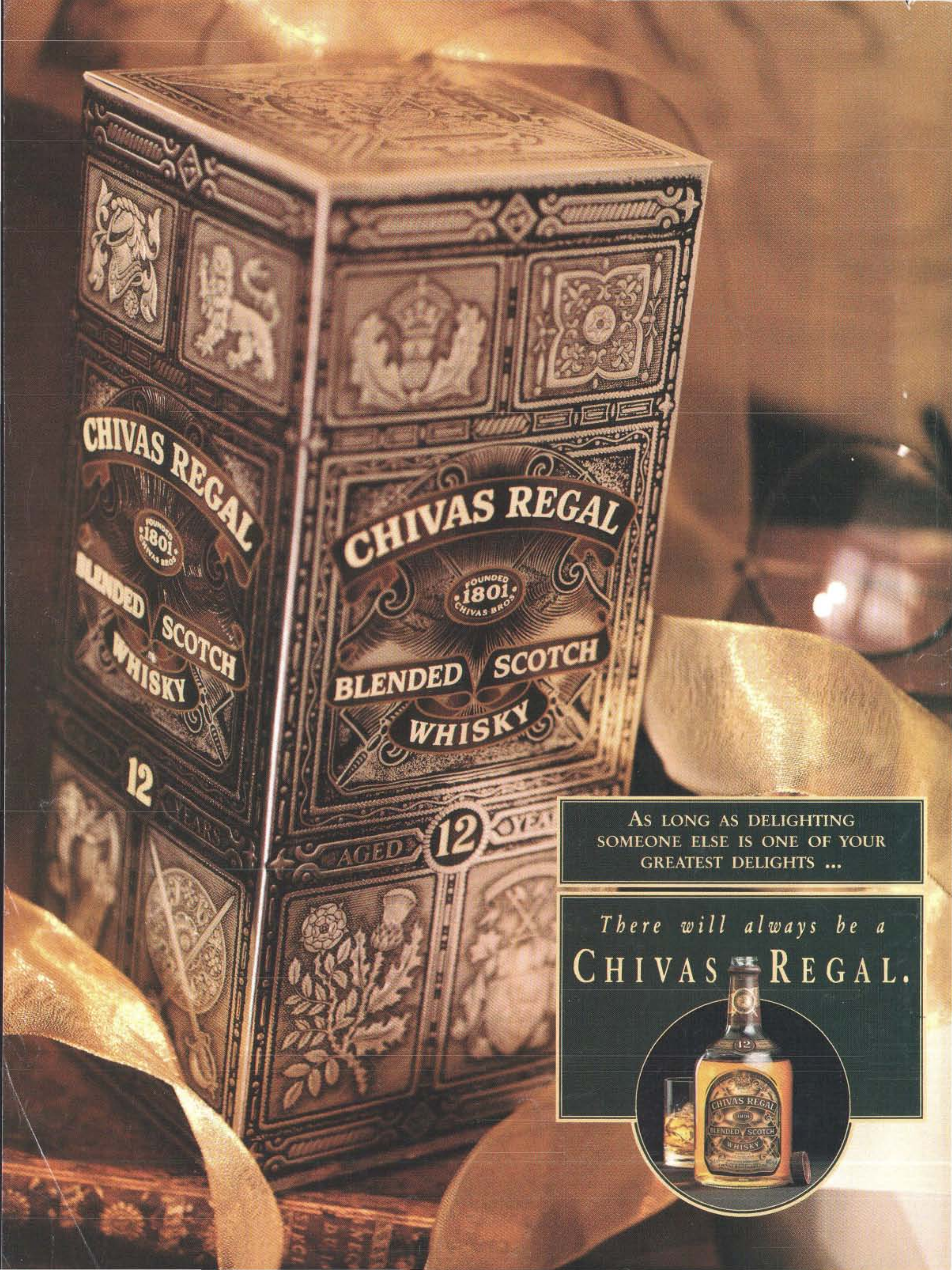
The breakthroughs we're pioneering today are transforming an entire generation of vital space and defense programs. In fact, Lockheed is involved in all of the key technology areas that the Department of Defense and the Defense Advanced Research Projects Agency deem critical for America's future.

Preeminence in R&D gives Lockheed advantages in aeronautics, missiles, space and military electronics. It makes us stronger competitors; better able to manage national priority programs and deliver technology that is both reliable and affordable. Just what you'd expect from a premier aerospace company. Just what you get from Lockheed.

Lockheed leads.

Lockheed

Watch NOVA on PBS, Tuesdays at 8 p.m.



CHIVAS REGAL
FOUNDED 1801 CHIVAS BROS
BLENDED SCOTCH WHISKY
AGED 12 YEARS
AS LONG AS DELIGHTING SOMEONE ELSE IS ONE OF YOUR GREATEST DELIGHTS ...
There will always be a
CHIVAS REGAL.